Determinants of clinical phenotype in myeloproliferative neoplasms

Jacob Grinfeld

Girton College

September 2018

This dissertation is submitted for the degree of

Doctor of Philosophy

Preface

This dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration, except as detailed in the Acknowledgements and where specified in the text.

It is not substantially the same as any that I have submitted, or, is being concurrently submitted for a degree or diploma or other qualification at the University of Cambridge or any other University. I further state that no substantial part of my dissertation has already been submitted, or, is being concurrently submitted for any such degree, diploma or other qualification at the University of Cambridge or any other University of similar institution.

This thesis does not exceed the word limit set by the Degree Committee for the faculty of Clinical Medicine and Veterinary Medicine.

Acknowledgements

Firstly, I would like to thank my supervisors, Professor Anthony Green for giving me the opportunity to carry out this work in his group and Dr Peter Campbell, for his helpful discussions and advice regarding statistical analysis and genomics, and for their supervision over this period. Jyoti Nangalia was instrumental to the success of this project. She designed the original project including bait set design, sample collection and submission for sequencing, together with Elli Papaemmanuil. I worked with her to filter the variants and independently validate clinical events and she worked with me in much of the down-stream analysis. I would like to thank her for all her guidance and advice over the years and also thank Elli for her support with initial variant analysis and advice in general. Joanna Baxter, and the team at CBSB, were also vital to the success of this project and were involved in DNA and serum sample acquisition and processing, as well as providing much of the clinical data used. I would like to thank the team at the Cancer Genome Project for carrying out the library preparation, sequencing and the Caveman and Pindel pipelines. Gunes Gundem helped with a number of the variant filtering algorithms. I would like to thank everyone in the Green lab for their help, friendship, teaching and constructive criticism over the years, in particular Fran Nice (who was involved in colony analyses and helped filter the DNMT3A variants), Carlos Gonzalez-Arias (for his help and support), David Flores Santa Cruz (who additionally assisted me with some of the bash and R coding), Thorsten Klampfl (for his fruitful statistical and genomics discussions) and Juan Li (for her teaching and advice).

The cytokine work done described in chapter 6 was done in conjunction with Dave Kent and his lab, namely Nina Friesgaard Oebro, Miriam Belmonte, and Olivia Harris and I am grateful for their helpful and fruitful collaboration. Dave has been source of helpful advice and discussion throughout the PhD. I would also like to thank Anna Godfrey, Cathy McLean, Julie Temple and Julia Cook for their help with the PT-1 datasets. I am grateful to Clare Harrison, Alessandro Vannucchi, Paola Guglielmelli, Christen Andersen and Hans Hasselbalch for sharing their samples and clinical data with us.

Nicos Angelopoulos carried out the Bayesian network analysis. David Wedge performed analysis of mutation order. The multi-stage modelling was carried out with Dr Moritz Gerstung and Dr Rob Cantrill, and I am grateful for their advice, discussion and valuable input. Finally I would like to give my thanks and love to my family: my parents, sister, wife Aurora and daughter Freya; particularly to Aurora who has put up with me and supported me through what has at times been a difficult period.

Determinants of Clinical Phenotype in Myeloproliferative Neoplasms

Dr Jacob Grinfeld

Background: Myeloproliferative neoplasms, (MPNs) such as polycythemia vera, essential thrombocythemia, and myelofibrosis, are chronic hematologic cancers with varied progression rates. The genomic characterization of patients with myeloproliferative neoplasms offers the potential for personalized diagnosis, risk stratification, and treatment.
Methods: We sequenced coding exons from 69 myeloid cancer genes, common to two separate bait sets, in patients with myeloproliferative neoplasms, comprehensively annotating driver mutations and copy-number changes or copy. We developed a genomic classification for myeloproliferative neoplasms and multistage prognostic models for predicting outcomes in individual patients. Classification and prognostic models were validated in an external cohort. Cytokine profiles of over 400 patients were also analysed to determine the contribution of the inflammatory microenvironment to phenotype and progression risk.
Results: A total of 2035 patients were included in the analysis. 33 genes had driver mutations in at least 5 patients, with mutations in *JAK2*, *CALR*, or *MPL* being the sole abnormality in 45% of the patients. The numbers of driver mutations increased with age and advanced disease. Driver mutations, germline polymorphisms, and demographic variables independently predicted whether patients received a diagnosis of essential thrombocythemia as compared with polycythemia vera or a diagnosis of chronic-phase disease as compared with myelofibrosis. In particular a set of mutations that included ASXL1, SRSF2, U2AF1 and EZH2 was enriched in myelofibrosis and associated with poor outcomes. The JAK2 46/1 haplotype strongly correlated with the presence of 9pUPD and independently with a PV phenotype, demonstrating that the underlying germline background can play a role in determining somatic events and can affect the patient's phenotype in its own right.
We defined eight subgroups based solely on clustering of genomic data that showed distinct clinical phenotypes, including blood counts, risk of leukemic transformation, and overall survival. These included a sub-group defined by mutations the same set of chromatin and splicesome component genes described above, and a subgroup enriched for TP53 mutations and chromosomal changes, which carried a significant risk of AML transformation. Patients with no detectable mutations had very low rates of progression or death. By integrating 63 clinical, demographic and genomic variables, we created prognostic models capable of generating personally tailored predictions of clinical outcomes in patients with chronic-phase myeloproliferative neoplasms and myelofibrosis. The predicted and observed outcomes correlated well in internal cross-validation of a training cohort and in an independent external cohort. The prognostic model performed as well as or better than a number of existing risk scores including the high molecular risk genetic score and international prognostic scoring systems and even within individual categories of existing prognostic schemas, our models substantially improved predictive accuracy.
Cytokine profiles varied significantly across MPN subtypes, with high levels of TNF-alpha and IP-10 seen in myelofibrosis, and to a lesser extent in polycythemia vera. Patients with essential thrombocytosis however, were found to have high levels of GRO-alpha and EGF, and levels of these at single time points or when measured longitudinally were predictive for the risk of progression to myelofibrosis.
Conclusions: Comprehensive genomic characterization identified distinct genetic subgroups and provided a classification of myeloproliferative neoplasms on the basis of causal biologic mechanisms. Integration of genomic data with clinical variables enabled the personalized predictions of patients' outcomes and may support the treatment of patients with myeloproliferative neoplasms.

Contents

1. Introduction

1.1 Philadelphia-negative Myeloproliferative neoplasms

The three classical Philadelphia-negative myeloproliferative neoplasms (MPNs) are characterised by clonal expansion of haematopoietic progenitors, hypersensitivity to, or independence from, cytokines, and overproduction of mature erythroid and myeloid progeny. Common to them are the clinical features of bone marrow hypercellularity, and an increased risk of thrombosis, haemorrhage and transformation to acute myeloid leukaemia. Since these are chronic conditions that normally manifest well in advance of leukemic transformation they offer an invaluable model for studying the process of leukaemogenesis and advancing our understanding of stem cell function and cell fate decisions. A better understanding of the processes that determine the progression to acute leukaemia are of particular importance in the context of MPNs as these leukaemias are invariably not amenable to standard chemotherapy.

Current diagnostic criteria separate the classical Philadelphia-negative MPNs into three distinct disease entities: polycythaemia vera – characterised by a raised red cell mass; essential thrombocytosis – characterised by an isolated increase in platelet count; and idiopathic/primary myelofibrosis – in which the haematopoietic compartment is gradually replaced with collagen fibres, leading to bone marrow failure and extramedullary haematopoiesis, and which is often associated with constitutional symptoms[1].

1.2 Evidence for clonality

Support for the clonal, and therefore neoplastic, nature of these conditions came from studies in the 1970s that found skewing of lyonisation at the G6PD locus (on the X chromosome) in women with PV[2] and ET[3]. Additionally, a number of clonal markers were found. These included deletions of chromosome 20q in a small number of MPN cases (predominantly PV), as well as in other myeloid malignancies[4,5], gains of chromosomes 8 and 9 and deletions of chromosomes 5, 7 and 13[6]. More striking was the discovery of frequent (~33% in early studies) copy-number neutral loss of heterozygosity on chromosome 9p, identified using microsatellite polymerase chain reaction (PCR) and missed on standard cytogenetic analyses[7].

1.3 JAK2 mutations

1.3.1 Discovery and characterisation of JAK2 mutations

In 1951, William Dameshek hypothesised that rather than these being "pure" proliferations, the three myeloproliferative phenotypes may represent differing manifestations of a single underlying process[8]. This was shown to be the case when mutations in components of cytokine signalling pathways were detected across different MPN phenotypes. The commonest of these is the valine to phenylalanine substitution at codon 617 (V617F, due to a G>T substitution), of the Janus kinase 2 (*JAK2*) gene on chromosome 9p (and located in the region of previously identified LOH) which is found in 95% of patients with PV and 50-60% of those with ET and PMF[9–12]. JAK2 is a cytoplasmic tyrosine kinase, required for signal transduction from a number of cytokine receptors, including those for thrombopoietin (Tpo), erythropoietin (Epo) and granulocyte colony stimulating factor (G-CSF), and which therefore plays a vital role in haematopoiesis.

1.3.2 Effect on molecular function

The V617F mutation occurs in the JH2 (janus homology 2 or "pseudokinase") domain and results in a constitutive increase in the activity of the JH1 kinase domain. The mechanism by which this occurs is not entirely clear, but it is thought that the mutation may reduce the autoinhibitory function of the JH2 domain via changes in JH2 conformation[13], ATP binding[14] and its own inhibitory kinase activity[15]. Although expression of *JAK2*V617F has been shown to allow for JAK2 signalling in the absence of receptor ligation[10,11], the expression of cytokine receptors[16] and a functional FERM domain (required for receptor binding)[17] are still required for JAK2 signalling and cytokine independent growth. Furthermore, there is also evidence that the V617F mutation allows escape from negative regulation by suppressor of cytokine signalling 3 (SOCS3)[18]. This increased JAK2 signalling recapitulates that seen in the physiological response to cytokine binding, namely the increased activation of signal transducer and activator of transcription (STAT) 5 and 3, mitogen associated protein (MAP) kinase and Phosphoinositide 3-kinase (PI3k) pathways[10,12]. STAT5 in particular is thought to have a critical role in the pathogenesis of PV, as its deletion abrogated a PV phenotype in a JAK2V617F-knockin mouse model[19]. Transcriptional changes seen in JAK2-mutated cells include upregulation of polycythemia rubra vera-1 (PRV-1)[20,21], the function of which has been associated with increased signalling in response to Tpo and interleukin-3 (IL-3), and of nuclear factor erythroid-2 (NFE2), a transcription factor implicated in erythropoiesis and megakaryopoiesis[22]. Over-expression of NFE2 even in cells lacking *JAK2* mutations has been associated with erythropoietin-independent erythroid maturation and red cell overproduction[23,24], and can

induce a myeloproliferative phenotype, with thrombocytosis and leucocytosis, in mouse models[25].

Mutations in exon 12 of JAK2 are also seen in a small proportion of patients and are located between F533 and F547, in the linker region between the SH2 and JH2 domains[26,27]. These mutations also induce increased constitutive signalling to a greater degree than *JAK2*V617F, with greater JAK2 and extracellular regulated kinases 1 and 2 (erk1 and erk2)[26], and cytokine independent growth. The mechanism by which these mutations act is less well understood.

In addition to its in cytoplasmic signalling, a number of non-canonical roles have been described for JAK2, including histone phosphorylation[28], which may be perturbed by pathogenic mutations and may provide alternative pathways for dysregulation of transcription and stem cell function. J*AK2*V617F expression has also been associated with an increase in reactive oxygen species, activation of DNA-repair mechanisms, disruption of cell-cycle checkpoint responses and impaired apoptotic response to DNA damage, and accordingly with a greater rate of double-strand breaks homologous recombination[29–31]. Therefore *JAK2* mutations appear to be able induce a mutator phenotype, which may be one mechanism underlying progression of disease to MF, MDS or AML.

1.3.3 Effect on cell biology

A number of studies have demonstrated that JAK2 mutation can be found in the haematopoietic stem cell and progenitor (HSPC) compartment[32–34]. Xenograft models suggest that *JAK2* mutations do not however result in a self-renewal advantage[34] and this is recapitulated in some knock-in mouse models, where JAK2 positive HSPC are skewed towards symmetrical differentiation and expansion of the progenitor pool rather than self-renewal, and demonstrate a lack of an advantage in competitive transplantation experiments[35,36]. This also is consistent with the absence of significant clonal expansion over time or development of overt MPN in cases with a transplanted *JAK2*-mutant clone[37] or in a subset of cases in the general population that have developed a detectable *JAK2*-mutant clone[38] (see later discussion). These findings have lead to the suggestion that the JAK2 mutations alone are insufficient to lead to persistent disease and that additional mutations, or other factors, are required to overcome the lack of a competitive advantage. It is also possible however expansion of the progenitor pool alone is sufficient to establish and sustain disease, and that a stem cell advantage is not required at all. This is consistent

with the finding that adult haematopoiesis is predominantly sustained by a pool of long-term multipotent progenitors[39].

1.3.4 Acquired 9pUPD and JAK2 homozygosity

As noted above, acquired uniparental disomy/loss of heterozygosity of chromosome 9p had been found to be a recurrent abnormality in patients with PV even before the discovery of *JAK2*V617F mutations. It is now evident that this genetic lesion is found almost exclusively in the context of *JAK2* mutations, and results in their homozygosity. Using fluorescence microsatellite PCR *JAK2* homozygous subclones were identified in up to 80% of PV patients and around 50% of *JAK2*-mutated ET patients, although the ratio of homozygous to heterozygous clones tended to be much higher in PV cases, with homozygous clone sizes generally <10% in patients with ET[40]. A link therefore exists between *JAK2*V617F dosage and the degree of erythrocytosis and this is supported by the finding of higher haemoglobin concentrations (as well as leucocyte counts) in JAK2-homozygous patients with ET[41,42]. This hypothesis is additionally supported by a number of experimental models where the ratio of wild-type to mutant alleles correlates with increasing independence from Epo and increasing erythropoiesis[36,43,44]. Similarly, JAK2 exon 12 mutations, associated with greater JAK2 and erk1/2 phosphorylation, are always associated with erythrocytosis and do not appear to result in thrombocytosis.

However, the presence of homozygosity does not account for the differences seen between patients with ET and PV alone, since, as noted here, not all patients with PV have homozygous clones, while conversely a number of patients with ET do. Furthermore, transcriptional differences are seen in erythroid colonies from ET and PV patients when analysis is restricted to heterozygous colonies alone[45].

1.4 Other intracellular signalling pathway mutations

1.4.1 MPL mutations

Missense mutations at codon 515 of *MPL* (which encodes myeloproliferative leukaemia protein, the thrombopoietin receptor itself), including M515L and M515K, are also seen and are associated with increased STAT-3, STAT-5, ERK and AKT signalling[46,47]. Less frequently, other activating mutations of MPL, such as S505N, which is also found as a germ-line variant in familial ET, and S204P are also seen and are associated with thrombopoietin hypersensitivity or independence[48–50]. Given *MPL* is involved in thrombopoietin signal transduction and megakaryocytic differentiation, it is not surprising that these mutations are seen almost exclusively in patients with ET and MF (roughly 8-

10% of *JAK2* unmutated cases), and not found in those with PV.

1.4.2 CALR mutations

More recently mutations in calreticulin (CALR) have been identified in patients with ET and MF. These have consistently found to be insertions or deletions (most commonly a 52bp deletion or 5bp insertion) in its final exon resulting in a 1bp shift in the reading frame and a common novel C-terminal sequence[51,52]. This was an unexpected and unusual result, since, unlike JAK2 or MPL, CALR is not known to have a direct role in megakaryopoiesis or regulation of haematopoietic stem cells. Rather, it is known to be involved in the regulation of calcium uptake and release by the endoplasmic reticulum (ER), and acts as a chaperone, regulating folding and quality control of newly synthesised glycoproteins in the ER[53,54]. However, the restriction of CALR mutations to patients with ET and MF, mutual exclusivity for JAK2 and MPL mutations, its preferential expression in megakaryocytes (as determined immunohistochemically)[55] and increased JAK-STAT signalling in *CALR* mutant cells[51,56], pointed towards a similar mechanism to *MPL* mutations.

Subsequently it was demonstrated that CALR mutations can impart TPO-independence in cell lines[57–59], retroviral and transgenic mouse models[60–62] in a MPL- and JAK2-dependent manner, mimicking the effect of activating MPL mutations and recapitulating their phenotype *in vivo*. This has been shown to be mediated by direct binding of MPL by the N domain of CALR, a phenomenon that uniquely occurs in the presence of the mutated C-terminus, leading to autocrine activation of MPL, JAK2 dimerization and downstream STAT5 and ERK phosphorylation[59–61].

Other mechanisms may still be in play, beyond aberrant MPL binding and activation. The loss of the KDEL (lysine, aspartic acid, glutamic acid, leucine) signal from the C-terminus may lead to mislocalisation of the protein and therefore to other aberrant functions. It has also been shown that type 1 (52 bp deletion), or type-1 like, and type 2 (5 bp deletions), and type 2-like, mutations are associated with different clinical phenotypes: Type 1-like mutations are associated more with MF and a higher rate of MF transformation when found in ET but a more favourable risk for patients with MF[63,64], while Type 2-like mutations associate more with ET and higher platelet counts[65]. This has associated with differences in calcium binding, with the most aberrant ER-dependent calcium shifts seen in with Type 1/type 1-like mutants, which are associated with loss of a greater number of

calcium binding sites[65].

1.4.3 Mutations in regulators of cytokine signalling (CBL, SH2B3, SOCS)

Compared to other malignancies, the MPNs have relatively few mutational events with the majority carrying mutations in JAK2, CALR or MPL (Figure 1A). However, a number of additional genes have been found to be mutated in up to 10-20% of cases (Figure 1B). These have been investigated both in smaller studies sequencing individual genes across myeloid malignancies as discussed in the following sections, or as part of wider targeted[66–69] or whole exome sequencing studies[52,70].

As might be expected from the role of receptor signalling pathways in the pathogenesis of MPNs, loss of function mutations in negative regulators of receptor tyrosine kinases are seen, as well as mutations in down-stream pathways.

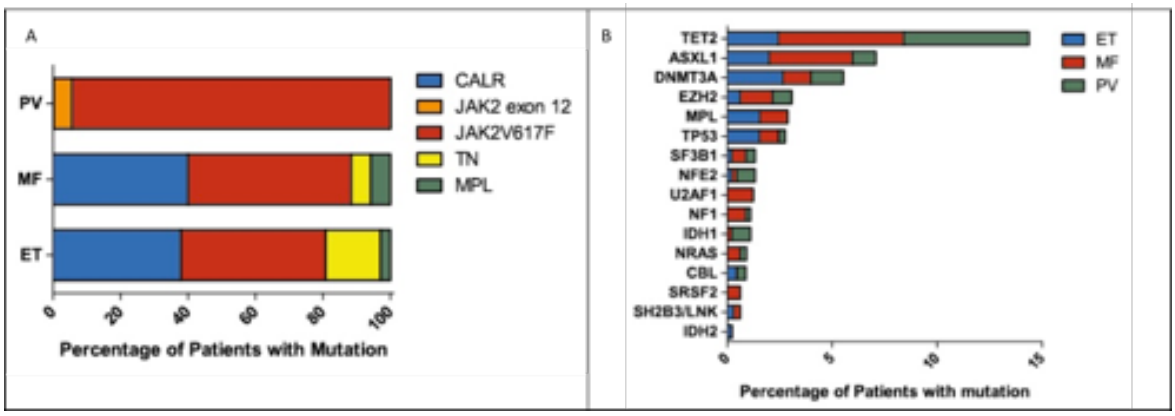


***Figure 1**: (A) Frequencies of JAK2, CALR and MPL mutations across MPNs. TN denotes "triple negative" (B) Frequencies of additional mutations, with counts normalised to MPN subtype frequency[52,66,95].*

LNK (lymphocyte specific adaptor protein, or SH2B3) is an adaptor protein that binds both MPL and JAK2 to act as a negative regulator of JAK-STAT signalling. LNK deficient mice display an MPN-like phenotype with megakaryocytic hyperplasia, cytokine hypersensitivity and splenomegaly[71]. LNK exon 2 mutations, affecting the plekstrin-binding domain, are found in a small number of MPN patients, more often seen in advanced phase disease, and can also be found in patients with an erythrocytosis lacking a JAK2 mutation, suggesting they may be sufficient to initiate disease[72–74].

Another negative regulator of cytokine signalling, Casitas B-cell lymphoma oncogene (CBL) is an E3-ubiquitin ligase, which is specifically involved in the ubiquitination (and resultant degradation) of a number of receptor tyrosine kinases, as well as having a role in

signal transduction itself. Its targets include PDGFR, c-KIT, FLT3 and MPL[75]. Mutations in its RING-domain (responsible for ligase activity) have been described in MPNs, particularly in MF, atypical chronic myeloid leukaemia (aCML) and chronic myelomonocytic leukaemia (CMML)[76,77]. Mutations in the SOCS proteins, which also suppress JAK2 signalling, are rarely seen in MPNs[78] but their down regulation has been observed[79].

1.4.4 Ras pathway: NRAS, KRAS, NF1, PTPN11

Mutations increasing RAS pathway signalling are also observed, but, as with CBL and LNK mutations, tend to be seen more in myelofibrosis, advanced/transformed disease or myeloproliferative neoplasm/myelodysplasia (MPN/MDS) overlap conditions[80]. In mouse models NRAS mutations are sufficient to initiate a CMML-like disease characterised by myeloid bias and increased self-renewal, but with a propensity for proliferative exhaustion[81]. KRAS mutations as well as those in negative regulators of RAS, such as NF1 and PTPN11, have also been found in some MPN patients[66,67,82].

1.5 Mutations in epigenetic regulators

1.5.1 TET2 mutations

Mutations in epigenetic regulators are the commonest mutations seen in addition to those affecting *JAK2*, *CALR* and *MPL*. In some cases, these mutations are enriched in patients with MF or MPN/MDS overlap conditions, but often are found across the spectrum of MPNs. Of these, loss-of-function mutations of ten-eleven translocation family member 2 (TET2) are the most common and are detected in approximately 10% of MPNs[83,84]. TET2 hydroxylates 5-methylcytosine to 5-hydroxymethylcytosine (5-hmc), a process which is thought to be particularly important for transcriptional regulation in stem cells and in embryonic development. A reduction in 5-hmc is observed in *TET2* mutated patients, and is associated with increased self-renewal capacity and myeloid bias[85–87]. The HOXA cluster, implicated in cell fate decisions, is known to be regulated by TET2, and therefore its dysregulation in the context of *TET2* mutations may be one mechanism underlying the stem cell expansion seen with these mutations[88].

1.5.2 DNMT3A mutations

Loss-of-function mutations (including dominant negative missense mutations at codon 882) in DNA methyltransferase 3a (DNMT3a), a protein responsible for de novo methylation of CpG dinucleotides, are found across MPN subtypes[89]. As with *TET2*

mutations, their exact role in MPN pathogenesis is not yet fully understood, but it is thought that the resultant epigenetic deregulation (both hypo- and hypermethylation are seen) results in upregulation of "HSC fingerprint" genes such as GATA3 and RUNX1 and downregulation of differentiation factors such as *Ikaros* (*IKZF1*), resulting in a differentiation block and HSC expansion[90,91].

1.5.3 IDH mutations

Hotspot mutations in isocitrate dehydrogenase 1 and 2 (*IDH1* and *IDH2*) are described in <5% of cases[66,92] These enzymes catalyze the conversion of isocitrate to alpha-ketoglutarate, but these mutations result in the production of 2-hydroxyglutarate, which inhibits Jumonji-C domain histone demethylases. This leads to histone hypermethylation and also inhibits TET2 activity, which in turn results in a differentiation block[93,94]. In contrast to TET2 mutations, which not appear to be associated with particular MPN subtypes (while being associated, in one study, with poorer overall survival and increased progression to AML[95]), IDH mutations are more commonly found in MF or transformed disease[92,96].

1.5.4 Mutations in polycomb repressor complex components

Mutations of genes involved in histone methylation are over-represented in myelofibrosis/transformed disease. EZH2 (PcG Enhancer of Zeste Homolog 2) is the catalytic component of the polycomb repressive complex 2 (PRC2) and, together with EED and SUZ12 acts to trimethylate histone H3 lysine 27 (H3K27) causing repression of transcription. In contrast to *EZH2* mutations seen in lymphoma, those in MPNs tend to be loss-of-function mutations[97,98] that result in de-repression of a number of genes (including putative oncogenes such as *LMO1* and *HOXA9*) and increased HSC self-renewal[99,100].

Mutations of Additional sex combs like 1 (ASXL1) are also relatively common in MF[66,101]. ASXL1, a component of PRC1, is also known to regulate PRC2 and to have a role in the regulation of HOX genes. ASXL1 mutations accordingly are associated with HOXA upregulation and reduction in H3K27 methylation, which were linked with impaired recruitment of EZH2[102].

1.5.5 Mutations in spliceosomal proteins

Mutations in components of the spliceosome including splicing factor 3B subunit 1 (SF3B1), splicing factor arginine/serine-rich 2 (SRSF2), U2 small nuclear RNA auxiliary

factor 1 (U2AF1) and zinc finger RNA-Binding Motif And serine/arginine rich 2 (ZRSR2) as well as other much rarer mutations, are also seen in MPNs, again more frequently in patients with MF of MPN/MDS overlap syndromes, and are often associated with poorer outcomes[66,103–105].

The effects of hotspot mutations of codon 95 of SRSF2, codon 34 mutations of U2AF1 and codon 700 of SF3B1 are perhaps the best explored, as knock-in mouse models have been made for each[106–108]. SRSF2 mutations appear to result specifically in skewed motif recognition (rather than loss of function) with alterations in exon usage in a number of genes, which include EZH2 (leading to reduced expression), and bcl-6 co-repressor (BCOR), which is also known to be mutated in myeloid malignancies[106]. Mutations in U2AF1 alter its 3' splice acceptor preferences leading to mis-splicing of a set of genes that includes BCOR and SRSF2[107]. EZH2 expression was also found to be reduced in 63% of U2AF1/SRSF2-mutated primary specimens (with a greater amounts of unspliced than spliced transcripts), and this was associated with decreased H3K27 trimethylation[100].

Spliceosomal mutations are common in MDS. Mutations of SF3B1 in particular have been associated with the presence of ring sideroblasts, with the specific syndrome of MDS with ring sideroblasts and thrombocytosis (RARS-T) and with a much better prognosis than other spliceosome component mutations, as well as being found in a proportion of patients with chronic lymphocytic leukaemia[109,110]. These mutations are particularly associated with the use of cryptic 3' splice sites and lead to the mis-splicing of a number of genes including those involved in mitochondrial metabolism and iron homeostasis, as well as *ASXL1* and *CBL*, resulting in defective erythropoiesis[108,111,112].

1.5.6 Mutations in other transcriptional regulators

Protein truncating insertion/deletion mutations have been reported in NFE2[113]. These mutations were predominantly restricted to patients with PV or post-PV MF, and were associated with both higher NFE2 mRNA levels and increased protein stability. As noted above, NFE2 is thought to be a critical regulator of megakaryopoiesis and erythropoiesis, with roles in the regulation of haem and β-globin synthesis, and its over-expression has been reported across MPN phenotypes, as well as in other conditions associated with erythropoietin hypersensitivity and erythrocytosis[114,115]. Mutations in RUNX1 which also plays an important role in the transcriptional regulation of stem cell fate decisions have also been reported in MPN, but associate more with leukaemic transformation of the

disease[116], while GATA1 mutations are seen in the context of transient myeloproliferative disorder[117]. *IKZF1*, which is commonly mutated in acute lymphoid leukaemia, are found to be deleted in a small number of MPN cases transformed to acute leukamia[118].
Finally, mutations in the cohesin complex components (including STAG2), which has roles in sister chromatin adhesion, transcriptional regulation and DNA repair, have been found in a small number of MDS/MPN overlap cases[119,120].

1.6 The impact of mutation acquisition order

While JAK2, CALR and MPL mutations are almost always mutually exclusive, the further mutations discussed above are normally found in combination with a phenotypic driver, and in combination with each other. There is increasing evidence that not only can the specific combinations present in an individual influence their disease phenotype and natural history, but the specific clonal architecture, including the order in which the mutations were acquired, and relative clone sizes present for each may do as well. This may in part explain the clinical heterogeneity across even patients with the same mutations or combinations of mutations[121].

While CALR mutations are shown in most cases to occur in the dominant clone, differences in the order of mutation acquisition have been demonstrated in JAK2-mutated MPNs harbouring concurrent TET2 or DNMT3A mutations, where mutation order has been shown to have an impact on disease phenotype, thrombotic risk, age of presentation and response to treatment[122,123]. Patients in whom the JAK2 mutation occurred first have larger “double mutant” subclones (harbouring both JAK2 and either TET2 or DNMT3A mutations), as well as larger JAK2-mutated homozygous clones. In JAK2-first TET2-second cases, there is expansion of erythoid progenitors, and patients present at a younger age and are at greater risk of thrombosis.

In contrast, TET2-first or DNMT3A-first patients are characterised by a dominant “single positive” subclone (i.e. harbouring the TET2/DNMT3A mutation only) in keeping with the greater self-renewal capacity seen with TET2/DNMT3A mutations compared to those in JAK2. These patients were more likely to have ET, which may in part be related to constraints on the expansion of the JAK2-mutated clone and on the development of homozygosity. Therefore, it is likely that the order in which mutations occur will influence the composition of the stem cell niche in which later sub-clones will arise and reside, potentially introducing constraints on their potential to expand. However, it is also feasible

that separate clones may act cooperatively. This has been shown in solid tumors such as glioblastoma, where cytokine secretion by a relatively minor subclone can drive the expansion of a more dominant tumor clone[124].

Beyond determining clonal hierarchical structure, there is also evidence to suggest that the order that mutations occur can have cell-intrinsic differences in terms of transcription and clonogenic potential. For example, acquisition of a JAK2 mutation on a wild-type background was shown to be associated with a proliferative advantage, but this was not the case on the background of an earlier TET2 mutation. Furthermore, TET2 and JAK2-mutated (double mutant) HSCs/progenitors from JAK2-first patients were able to generate more progenitors than those from TET2-first patients, and this was reflected in differences in gene expression[122].

1.7 Additional Chromosomal changes

As discussed above, numerous recurrent cytogenetic abnormalities have been found in MPN patients through the use of conventional cytogenetic methods. Further studies have investigated their occurrence more systematically using high resolution single nucleotide polymorphism (SNP) microarrays or next generation sequencing (NGS) across MPN phenotypes in both chronic and transformed disease states, and enabling the identification of CN-LOH[125–128].

CN-LOH/UPD of chromosomes 1p and 9p (associated with and leading to homozygosity of *MPL* and *JAK2* respectively) are commonly seen, and associated with progression to myelofibrosis[125,126]. UPD of 4q, 7q, 11q, 14q and more rarely 8q are also seen. In some cases, these are known to be associated with mutated genes at these loci, leading to homozygosity for the mutant allele (including TET2 on 4q, *CUX1* and *EZH2* on 7q and CBL on 11q). However, this is not the case for all regions affected by UPD and it has been hypothesised that UPD serves to cause homozygosity for germline variants, for example *FANCM* R658X (located on chromosome 14q), acquired homozygosity of which was associated with transformation to leukaemia[129].

1.8 Impact of germ-line variation

The effect of inherited genetic variation on myeloproliferative phenotype forms a spectrum, from hereditary erythrocytosis or thrombocytosis, non-neoplastic conditions

where germline variants are sufficient give rise to myeloproliferation in all cases, to variants that increase an individual's risk of acquiring an MPN though the acquisition of classical MPN-associated variants or those that do not necessarily alter the risk of MPN development but may influence its phenotype. Additionally, genome wide association studies (GWAS) have identified a number of loci that are associated with haematological parameters such as haemoglobin concentration or platelet count[130], which may therefore influence the phenotype or presentation of an underlying MPN.

1.8.1 Hereditary erythrocytosis/thrombocytosis

The hereditary erythrocytoses/thrombocytoses are highly penetrant, monogenic disorders with Mendelian inheritance. Secondary hereditary erythrocytoses are driven either by mutations that give rise to high oxygen affinity haemoglobin variants, or mutations in oxygen sensing machinery, and are therefore associated with normal or high erythropoietin levels, while primary erythrocytoses are caused predominantly by Epo-receptor truncating mutations. Hereditary thrombocytoses are driven by mutations in the thrombopoietin gene itself (*THPO),* causing an increase in circulating levels, and in *MPL* as well as *JAK2.*

While these hereditary conditions are not MPNs, since the driving genomic variant is not somatically acquired and is present in all cells with polyclonal haematopoiesis, an understanding of the genetic basis for these conditions may still inform our understanding of MPNs. Furthermore, patients presenting with an apparent MPN may in fact have one of these conditions, and progression to bone marrow fibrosis and even AML is reported in cases with germ-line MPL mutations[131,132]. Additionally, the inherited variants seen here may also occur somatically, giving rise to an MPN. Examples of this include MPL S505N and JAK2 V617I mutations[48,133], but somatic mutations in other genes affected in these hereditary conditions may also have the potential to cause MPNs and should be considered in patients not found to have canonical JAK2, CALR or MPL mutations (triple negative cases).

The germ-line mutations of JAK2 are of particular interest, since they appear to give rise to over-production of a single lineage, for example isolated thrombocytosis is seen with JAK2 V617I[134] and erythrocytosis with E846D and R1063H[135], and secondly they do not appear to be associated with disease progression to bone marrow fibrosis or AML, highlighting important biological differences between these mutations and JAK2 V617F.

1.8.2 Familial MPN and germ-line predisposition

There are several reports of familial clustering of MPNs (where two or more family members are affected by MPNs), and there is a tendency (in approximately 60% of familial cases) for family members to present with the same MPN phenotypes. In fact, the relative risk of acquiring ET has been estimated to be approximately 12 times higher in first-degree relatives of MPN patients[136].

Causes of familial MPNs include germ-line mutations of RBBP6[137], and a high penetrance duplication of 14q32.2, which has been associated with overexpression of ATG2B, a mediator of autophagy, and GSKIP, a regulator of the WNT/β-catenin pathway[138]. It appears that these changes operate via independent pathways: RBBP6 mutations affect the p53 pathway and thereby influence the response to apoptotic stimuli and the risk of developing further mutations, while ATG2B and GSKIP overexpression promote megakaryopoiesis via increased thrombopoietin sensitivity.
A number of more common, but lower penetrance germ-line variants have been associated with MPN development, including single nucleotide polymorphisms (SNPs) present in, or close to, telomerase reverse transcriptase (TERT)[139] and "MDS1 and EVI1 complex locus" (MECOM)[140]. Two JAK2 haplotypes (46 and 1) are found to be in linkage disequilibrium (with the exception of one SNP). Whilst the combined haplotype (termed 46/1, or GGCC in reference to the defining alleles) is found in 24% of the population, it is found in up to 56% of patients with MPNs, with an odds ratio of 3 to 4[141–143]. Together with TERT and MECOM-associated SNPs, the 46/1 haplotype is estimated to account for 55% of the population attributable risk of developing an MPN[140].

Not only are JAK2 mutations more prevalent in patients with the 46/1 haplotype, the JAK2 mutation preferentially occurs on that allele in 46/1 heterozygotes, One possibility for this association is that the 46/1 haplotype is more prone to mutation and therefore more likely to give rise to JAK2V617F and exon 12 mutations. However, perhaps a more likely possibility is that the occurrence of mutant JAK2 in the context of the JAK2 46/1 haplotype confers an additional clonal or phenotypic advantage – the 'fertile ground' hypothesis. Further support for a clonal advantage for JAK2 46/1 is the fact that JAK2 allele burden rises more rapidly in patients carrying this also haplotype[144,145] but also that an association appears to exist between 46/1 haplotype and MPL mutations[146].

One other SNP, in the inter-genic region between HBS1L and MYB (rs9376092) has been

found to be enriched in MPL- and CALR-mutated MPNs, and more frequently in JAK2-mutated ET patients[140]. Together with the 14q32.2 duplication mentioned above, which was found to be predominantly associated with an ET phenotype, the HBS1L-MYB SNP appears to be one of the few germ-line variants that is associated with a particular MPN phenotype. Other reported genomic variants that may predispose more towards PV include SNPs within JAK2 itself[147] and affecting the glucocorticoid receptor (GR)[148,149] , although the latter association was not validated in a subsequent study[150].

Finally, it is worth noting that gender also appears to have an effect on the risk of MPN development and MPN subtype, with female predominance in ET and in younger patients with PV, but worse outcomes in male patients, although it is unclear to what extent this associated with intrinsic differences in tumour cell biology, or whether this relates to cell extrinsic factors such as levels of sex-related hormones, inflammatory cytokines or iron stores, or other potentially confounding factors[151].

1.9 Age-related clonal haematopoiesis

There has been increasing recognition that a proportion of individuals who do not have evidence of a myeloid malignancy still carry mutations in genes that are recurrently mutated in MDS, AML and MPNs.

Two early studies demonstrated that JAK2 V617F mutations were detectable in a proportion of healthy individuals: 5 of 57 healthy volunteers, and 36 of 3,935 patients in an unselected hospital population, roughly 10% in each case, although not necessarily representative of the wider population[152,153]. This was later confirmed, although at a lower rate, in a large study of the general population in Denmark where JAK2 mutations were found in 68 of 49,488 cases (0.1%), of whom 30 were already diagnosed with a myeloproliferative neoplasm. Of the remaining 33 cases, 18 developed an MPN within the subsequent 4 to 9 years[38]. Conversely, of a population of patients with a known MPN with previously stored samples, JAK2 mutations were present up to 15 years prior to the diagnosis of an overt MPN[145].

A set of recent studies have identified “clonal haematopoiesis” (i.e. a detectable clone carrying a mutation in a known recurrently mutated gene) as a common event in otherwise apparently healthy individuals, as has been labelled age-related clonal haematopoiesis or clonal haematopoiesis of indeterminate potential (ARCH or CHIP)[154–157]. The prevalence

of this phenomenon increases with age, with up to 10% of individuals over 60 carrying a detectable clone, which rose to over 20% in patients over 90. However, this was estimated at as high as 95% in one subsequent study of 50-60 year olds using more sensitive detection methods to detect clones with variant allele frequencies of as low as 0.0003[158]. The range of mutations detected was similar to that seen in MDS, AML and MPNs, with DNMT3A, TET2, JAK2, and ASXL1 mutated most frequently, but mutations in SF3B1, SRSF2, PPM1D, KRAS, NRAS, TP53, GNAS, CBL, GNB1, and STAG2 also seen[154–158]. Interestingly, mutations in spliceosome components were seen only in patients over 70[156] and mutations in CALR and MPL were not seen at all (although a number of these studies were geared towards only detecting single nucleotide substitutions and would therefore miss CALR mutations). The risk of subsequent progression to haematological malignancy was assessed in two of these studies. Although it was increased relative to the population without clonal haematopoiesis (hazards ratio 12.9 in one study), the rates of progression were less than 5-6% over the 6-10 year periods assessed[154,155].

Overall, these findings suggest that JAK2 mutations, among other MPN-related mutations, occur frequently in the general population without progression, in most cases, to an overt MPN, and when they do progress can do so after a long latent period. Possible explanations for this observation may include the requirement for subsequent, or pre-existing, somatic mutations or chromosomal changes, or an influence of the constitutional genomic background on which the JAK2 mutation occurs. However, previous sequencing studies suggest JAK2 mutations can often occur in isolation in MPN cases[52,66,95], and germ-line variants associated with JAK2 mutations in the context of MPNs appear to be widely similar to those predisposing for JAK2 mutation in the context of clonal haematopoiesis alone[159,160].

However, the competitive advantage for a potential clone depends not just on the mutation(s) that it carries and the germ-line background on which they occur, but also on the make-up of their competitors and on the environment in which they are competing. This is suggested by the finding of spliceosome complex mutations only in patients >70 years old, where it appears likely that they carry more of a competitive advantage in the context of the ageing marrow. This may be because of a loss of proliferative capacity in the wild-type stem cells present in the bone marrow, or differential responsiveness to micro-environmental signals such as increased IL-6 or IL-1beta or reduced CXCL-12 that may occur normally with age, or may be driven directly by the aberrant clone[161].

1.10 Role of microenvironment in MPN

A role for the bone marrow microenvironment in supporting the development of an MPN is evidenced by mouse models in which the deletion of Mib1 (causing dysregulated Notch signalling)[162] or of retinoic acid receptor-γ[163] solely in non-hematopoetic cells was sufficient to induce a myeloproliferative phenotype, suggesting that in some cases the clonal composition of the bone marrow itself may not be as important as the bone marrow environment.

Tumor necrosis factor α (TNFα), interleukin-6 (IL-6), fibroblast growth factor (FGF), interferon-γ inducible protein 10 (IP-10) and TGF-beta (TGF-β) production by bone marrow stroma, and potentially by the tumour clone itself, has been shown to promote the growth of MPN clones, while, in some cases inhibiting the growth of wild-type clones[164,165,7]. Tumour cells can also subvert their microenvironment. Clonal megakaryocytes and monocytes themselves (in mouse models, and from patients with myelofibrosis) secrete a number of cytokines, which include FGF, interleukin-8, TGF-β and vascular endothelial growth factor, that stimulate angiogenesis and drive fibroblast differentiation and recruitment, leading to bone marrow fibrosis[166,167]. In mouse models, JAK2-mutated clones have also been found to secrete lipocalin-2, which has been shown to suppress normal hematopoiesis via paracrine oxidative DNA damage, and may also drive the development of additional mutations in the tumor clone[168]. In addition, there is also evidence that, through direct cell-cell interactions and the secretion of soluble mediators such as TPO, C-C ligand 3 (CCL3) and interleukin-1β, the mutant clone can remodel the bone marrow niche to create an environment more permissive for its expansion, via depletion of sympathetic nerve fibres and nestin-positive mesenchymal cells[169] and expansion of osteoblast lineage cells[170].

However, despite an abundance of literature positing a role for micro-environmental inflammatory signals in the development of MPNs (including mathematical modelling)[171–174], and the in vitro work and mouse models mentioned above, data from patients with MPN are limited and occasionally contradictory, in part because the cytokines measured, methodology and choice of MPN subtypes vary considerably between studies.

Pro-inflammatory cytokines are consistently shown to be increased in MF[175–177], including IL-2, IL-6, IL-8, TNF-alpha, TIMP1, MIB-1b. However, the cytokine profile of patients with ET and PV is less well established and some studies showing greater increases in pro-

inflammatory cytokines in ET than PV[178], PV than ET[179] or showing no clear increase in ET[180]. It remains to be seen, therefore, to what extent cytokines contribute to the development and maintenance of MPNs, and whether differences between patients reflect differences in MPN phenotype.

1.11 Prognostic modelling in MPNs

While current practice aims predominantly to reduce the risk of thrombosis in patients with ET or PV, patients with myelofibrosis deemed to have a poor prognosis (in terms of overall survival or risk of leukemic transformation), can be treated with allogeneic bone marrow transplantation, which offers a potential cure. This decision currently is based on the use of prognostic scoring systems which, for the most part, utilise clinical parameters such as age, the presence of constitutional symptoms and blood counts. However, these scoring systems only offer categorisation into reasonably broad risk groups, and may be biased towards older patients with more advanced disease, who may be less fit to undergo transplantation.

Paired analyses have demonstrated the acquisition of a number of mutations or chromosomal changes, among others affecting TP53, IDH1/2, TET2, EZH2, CUX1 and IKZF1, suggesting a tumour suppressor role in the transformation process for these genes[94,104,125]. An association has also been shown between a number of mutations identified earlier in the natural history of the myelofibrosis, and subsequent shortened survival or leukaemic transformation, including mutations of ASXL1, EZH2, IDH1/2 and SRSF2[69,96,98,103]. Conversely, mutations in CALR have been associated with improved prognosis[63-65]. As a result, recent studies (discussed further in sections **5.1-5.2**) have started to integrate these genomic variables into the prognostic scoring systems used in MF. This may therefore improve our ability to better identify candidates who will benefit from bone marrow transplantation and offer the potential to do so earlier on the course of the disease.

Literature on predictive genomic factors in ET and PV is more sparse but the adverse factors identified show some overlap with those identified in myelofibrosis[68]. If predictive genomic factors can also be identified in patients with chronic phase disease, they may may identify high risk individuals within this patient group who might also benefit from transplantation or novel targeted therapies.

1.12 Questions

A number of questions still remain unanswered in this field.

- Are the phenotypic driver mutations sufficient to initiate and sustain disease, given that they do not appear to provide a stem cell advantage? If not, which additional factors (additional mutations, chromosomal events or micro-environmental factors) permit clonal expansion and disease maintenance?

- Do genomic changes underlie the ET or MF phenotypes seen in triple negative (TN) patients? If so, are these in known recurrently mutated genes or are novel genes affected?

- Are there distinct patterns of co-mutation and mutual exclusivity for genomic events, and from this can genomically defined diagnostic entities be determined?

- To what extent do additional mutations/chromosomal changes (beyond those affecting the phenotypic driver genes) determine phenotype? Are there additional factors determining phenotype (e.g. ET vs PV in JAK2 mutated patients)?

- To what extent can a wider and more comprehensive genomic characterisation of MPNs improve out ability to understand the mechanisms underlying disease progression and generate predictive models? Can these models be used successfully for chronic phase patients as well as those with myelofibrosis?

- What is the role of micro-environmental signals in determining disease phenotype and risk of disease progression?

1.13 Aims

- To comprehensively identify somatic variants and chromosomal changes and genotype SNPs associated with MPN pathogenesis or haematological parameters across a large MPN cohort to identify phenotypic correlates and genomic subgroups.

- To use demographic, clinical and genomic data to generate predictive models in order to identify those variables most likely to contribute to disease progression or

poor outcomes in MPN and to create clinically applicable prognostic models.

- To comprehensively measure cytokine levels in patients with MPNs and healthy controls to determine which cytokines might play a role in MPN pathogenesis and in determining phenotype and risk of disease progression.

2. Methods

2.1 Patient selection and sample processing

2.1.1 Patient cohorts

Peripheral blood samples from 1103 patients entered into the PT-1 trial were obtained. PT-1 is a multi-centre international trial where both newly diagnosed and previously treated patients with ET, aged 18 years or over, were recruited into one of three studies depending on their risk of vascular complications[181,182]. High-risk patients were randomised to receive either hydroxycarbamide plus aspirin or anagrelide plus aspirin[181], patients aged 40-60 years without previous thrombosis or cardiovascular risk factors entered the intermediate-risk study wherein they received either hydroxycarbamide plus aspirin or aspirin alone[182], and patients aged 40 years or under with no vascular risk factors entered the low-risk observational study and received aspirin only. All centres had appropriate research and ethical approval and patients gave their written informed consent. Clinical and laboratory details at and preceding diagnosis were obtained at trial entry, and patients were followed-up annually for treatment received, clinical events and blood counts. Patients fulfilled the current British Committee for Standards in Haematology (BCSH) criteria for ET[183].

Samples were additionally acquired from 21 ET, 33 PV, 16 MF, 2 MPNu patients enrolled in two phase II trials assessing the use of vorinostat[184,185], and from clinics at Addenbrooke's Hospital, Cambridge (n=231), Guys and St Thomas' Hospital, London (n=136) and Universita` di Firenze, Florence (n=366). The acquisition, storage and analysis of these samples was covered under the clauses of the 'Causes of Clonal Haematological Disorders Project' which had regional ethical approval from the Eastern Multi-region Ethics Committee (MREC 02/5/22 and 07/MRE05/44) and local research and ethical approval at participating UK hospitals.

2.1.2 Clinical data acquisition and curation

All patients in the study had their diagnoses made locally following the integration and review of clinical, laboratory and histopathological information at a multidisciplinary team, comprising the clinicians, specialist haemato-oncology histopathologists, and the laboratory molecular diagnostic team. Diagnostic criteria recommended by the British Committee of Standards in Haematology[183,186–188] and WHO[1] were followed for UK and

Italian patients respectively. PT-1 study patients met older Polycythemia Vera Study Group criteria for ET and despite criteria for ET having been revised multiple times over the course of the trial, central pathology review demonstrated high levels of concordance with current diagnostic criteria. The diagnostic review process for patients enrolled in the Vorinostat clinical study were as previously published[184,185]. Overall, this cohort comprises a mix of patients diagnosed in keeping with real-world best clinical practice.

Clinical and laboratory details at and preceding diagnosis were obtained in addition to follow up information on adverse events. These included age at diagnosis and date of diagnosis, sex, MPN subtype, prior history of thrombosis, treatment (where available), diagnostic blood counts (haemoglobin, white cell count, platelets), date of transformation to myelofibrosis, AML or death, presence of splenomegaly, and cytogenetics and/or reticulin grade (for patients who underwent bone marrow examination). These data were obtained from a number of sources including patient paper and electronic records (on EPIC, eMR and Meditech systems at Addenbrooke's Hospital), databases maintained by Cambridge Blood and Stem Cell Biobank (CBSB, Cambridge), the PT-1 trials database or directly from collaborators. Anonymised databases (containing unique patient identifiers) containing this data were maintained, and data formats standardised to enable further analysis.

Median follow-up was 107, 78 and 53 months for patients with a diagnosis of ET, PV and MF from the time of sampling respectively. The median time between diagnosis and sample acquisition was 55 days. In addition, 42% of patients were sampled within a month of diagnosis and 69% of patients were sampled within the first year from diagnosis.

Clinical and genomic data from an external cohort of 515 patients from the University of Florence Careggi Hospital were obtained. Diagnoses were again built up as part of an integrated clinical/laboratory/histopathology process, and each diagnosis was seen and discussed by 3 clinicians, lab biologists and the reference histopathologist, and if needed the histopathologist performed a second round of histological evaluation based on the clinical updated information.

2.1.3 Isolation of mononuclear cells from peripheral blood

40-60ml of venous blood was obtained by venepuncture and collected into lithium heparinized tubes and diluted 1:1 volume with phosphate buffered saline (PBS) at room temperature. The sample was then layered over an equal volume of a sodium

diatrizoate/polysaccharide density gradient (Lymphoprep; Axis-Shield, Dundee, UK) using a pipette on the lowest speed setting. Tubes were centrifuged at 800g at room temperature with no active deceleration. Mononuclear cells (MNC) were collected from the Lymphoprep/plasma interface using a pipette and washed twice in PBS. MNCs underwent red cell lysis if they were visibly contaminated with red cells - cells were incubated with 10ml of red cell lysis buffer (0.15 M NH4CL, 10mM NaHCO3, 0.1 mM EDTA in water) on ice for 5 minutes followed by addition of 10ml PBS and centrifugation at 300 g for 5 minutes to wash the cells of lysis buffer and lysed red cells. Cell pellets were resuspended in 1 ml of PBS and a cell count performed on a Woodley VetABC blood counter (Woodley, Bolton, UK) following the manufacturers instructions.

2.1.4 DNA extraction

20µl of RLT was mixed with double the volume of isopropanol and left for 15 minutes at room temperature before being centrifuged at 3600rpm for 45 minutes. The supernatant was removed by plate inversion and then washed with 25µl of 70% ethanol followed by centrifugation at 4500rpm for 10 minutes. The supernatant was then removed and the plate spun for <30 seconds to dry. DNA was then reconstituted in 40µl of dH2O.

2.2 DNA sequencing

2.2.1 Sanger sequencing

Primers for JAK2, CALR, PPM1D(A+B) and TET2 (specific for patient AQ31) were already available in the laboratory. PPM1D C+D primer sets were designed (using Primer 3 software) as the A+B primer sets were found to inadequately cover the area of interest. Primers were designed to be 18-22 base-pairs (bp) in length, have an annealing temperature of 60°C, and generate amplicons of 150-220bp in length. Primer sequences are shown in Table 1.

The PCR reactions used 6µl KAPA2G Fast ReddyMix TDS (Kapa Biosystems, MA, USA), 3.88µl water, 0.06µl 10nM forward primer, 0.06µl 10nM reverse primer and 2µl DNA to make a 12µl reaction volume. PCR conditions were as follows: 15 minutes at 95°C, followed by 32 cycles of 30 seconds at 95°C (denaturation step), 30 seconds at 56-62°C (annealing step) and 1 minute at 72°C (extension step), followed by a final extension step for 15 minutes at 72°C. PCR bands were visualized on an ethidium bromide containing 1.5% agarose gel.

Table 1: *Primers used for Sanger sequencing*

Primer Name	Forward primer sequence	Reverse primer sequence
JAK2V617F	TTTCCTTAGTCTTTCTTTG	TAGTTTACACTGACACCTAGCTGTG
CALR	CCTGCAGGCAGCAGAGA	ACAGAGACATTATTTGGCGCG
PPM1D_A	TGCCATCCTACTAGCTTCA	TTGGTCCATGACAGTGTTTGTTG
PPM1D_B	TTCCATTGGCCTTGTGCCT	AAAAAAGTTCAACATCGGCACCA
PPM1D_C	AATTAGTGAATGCATACCC	CAGAGTTCTTTCGCTGTGAGG
PPM1D_D	TGGCCTTTGTGCCTACTAC	TTTGATTTCTTTAAACATTTAGCCC
TET2(AQ31)	AATCCCATGAACCCTTAC	GGGTCTTGGCTTGGATACCT

The PCR reaction was first cleaned of excess dNTPs and oligonucleotides by mixing 5µl of PCR product with 0.08µl of exonuclease I (USB, CA, USA), 0.3µl of shrimp alkaline phosphatase (New England Biolabs, MA, USA), and 1.62µl of water. This mixture was heated at 30°C for 30 minutes then at 80°C for 15 minutes. Next, 2µl of the reaction product was mixed with 2.5µl of BDT buffer, 1µl of BDT, 1µl of 10nM primer and 3.5µl of water for a final reaction volume of 10µl. PCR cycling conditions were as follows: 1 minute at 96°C followed by 35 cycles of 10 seconds at 95°C, 5 seconds at 50°C and 4 minutes at 60°C. The PCR was run using both forward and reverse primers as separate reactions. Clean-up of the sequencing product was performed using by centrifuging using Cephadex gel. The supernatant was removed once again and the plate air-dried. 10µl formamide (Applied Biosystems, CA, USA) was added to each well prior to undergoing Sanger sequencing on a 3730xl DNA analyser (Applied Biosystems, CA, USA). PCRs performed at the Cancer Genome Project (CGP) laboratory utilised the same primers as above, however, reaction reagents and PCR conditions were different and followed CGP in-house protocols.

2.2.2 RNA bait set design

Two custom capture RNA bait libraries (Agilent SureSelect) were designed for targeted gene sequencing by Dr Jyoti Nangalia and Dr Elli Papaemmanuil and are here referred to as TGS1 and TGS2. These were chosen to provide coverage for genes known to be mutated in myeloid malignancies[189], as well as genes that were found to be recurrently mutated in a previous whole exome sequencing study in MPN[52]. The genes targeted by the

TGS1 bait set are show in Table 2, and by the TGS2 bait set in Table 3. In addition, 1966 SNPs were targeted in order to (a) cover SNPs that were known to be associated with an increased risk of MPN acquisition or had been reported in previous genome wide association studies (GWAS) to be associated with red cell or platelet indices in the normal population and (b) enable identification of genome-wide chromosomal changes. Sequencing data from these SNPs were not available for the 151 patients whose samples underwent exome sequencing rather than panel sequencing.

Table 2: *Genes targeted by the RNA bait set for TGS1. Overlap with TGS2 in bold.*

ABCC9	*CEACAM 6*	*EPHA7*	*IFNGR1*	*MSH4*	*PTPN4*	*SYNE2*
AC010872. 2	***CEBPA***	*ERVWE1*	*IL12B*	***NCL***	*PTPRA*	*TCF4*
ADAM18	*CELF1*	***ETV6***	*IL12RB1*	***NF1***	*PTPRB*	***TET2***
AHNAK	***CHEK2***	***EYA2***	*IL31RA*	***NFE2***	*PTPRC*	*TG*
AHNAK2	*CMYA5*	***EZH2***	*IL6ST*	*NFKB2*	*PTPRR*	***TP53***
AKAP9	*COL1A1*	***FAM47C***	***IRF1***	***NOTCH2***	*PTPRT*	*TRA2B*
ARHGAP32	*COL21A1*	*FANCA*	*ITPR1*	***NPM1***	***RAD21***	*TRPM4*
ASPM	*COL22A1*	***FARS2***	***JAK2***	*NR3C1*	***RAD51***	*TYRO3*
ASXL1	***CREBBP***	*FAT2*	*KCNMA1*	***NRAS***	***RB1***	***U2AF1***
ASXL3	*CSMD1*	*FAT4*	***KDM6A***	*NRD1*	*RBM14*	*U2AF2*
ATM	*CSMD3*	***FLT3***	*KIAA132 4*	*OCA2*	*RP1L1*	*VPS72*
ATRX	*CTNNA1*	***GABRB3***	*KIAA137 7*	*OTOP1*	***RUNX1***	***WT1***
BCLAF1	*CTNNA2*	***GATA2***	*KIAA154 9*	***PCDH15***	***SARDH***	*WWOX*
BCOR	*CTNND2*	***GNAS***	*KIF1B*	*PCDH7*	*SEZ6L2*	*ZBP1*
BRAF	***CUX1***	*GPR112*	***KIT***	*PDZRN4*	***SF3B1***	*ZBTB33*
BRD1	*DICER1*	*GPR116*	***KRAS***	***PHF6***	***SRSF2***	*ZEB2*
BRD9	*DNAH10*	*GRIN2A*	***KSR2***	***PPM1D***	*SGK2*	*ZFHX4*
BTNL8	*DNAH9*	***GRIN2B***	*L3MBTL 1*	*PRAGMIN*	***SH2B3***	*ZNF717*
C2orf39	***DNMT3A***	*HNRNPCL1*	***MBD1***	***PRKACB***	*SH3RF3*	*ZNF75D*
CACNA1G	***DNMT3B***	*HUWE1*	*ME1*	***PTEN***	*SLC12A1*	***ZRSR2***
CACNA2D 3	*DST*	*HYDIN2*	***MLL***	***PTPN11***	*SPAG9*	
CADM2	*DTNA*	***IDH1***	***MLL2***	*PTPN14*	*SRRM2*	

CALR	***ELF1***	***IDH2***	***MLL3***	*PTPN2*	*SSPO*	
CBL	***EP300***	*IFNA10*	***MLL5***	*PTPN21*	***STAG2***	
CDKN2A	*EPHA2*	*IFNAR1*	***MPL***	*PTPN3*	*SVEP1*	

Table 3: *Genes targeted by the RNA bait set for TGS2. Overlap with TGS1 in bold.*

ASXL1	***CHEK2***	***FAM47C***	***IRF1***	***NCL***	***PRKACB***	*SPTA1*
ASXL2	***CREBBP***	***FARS2***	***JAK2***	*NF1*	***PTEN***	***STAG2***
ASXL3	*CSF2RB*	*FBXW7*	***KDM6A***	***NFE2***	***PTPN11***	*SUZ12*
ATRX	*CSF3R*	***FLT3***	***KIT***	***NOTCH1***	***RAD21***	***TET2***
BCOR	***CTCF***	***GABR3***	***KRAS***	***NOTCH2***	***RAD51***	***TP53***
BRAF	***CUX1***	*GATA1*	***KSR2***	*NPM1*	***RB1***	***U2AF1***
C22ORF30	*DNMT1*	***GATA2***	***MBD1***	***NRAS***	***RUNX1***	***WT1***
C2ORF39	***DNMT3A***	***GNAS***	***MLL***	*OLFM4*	***SARDH***	***WWOX***
CACNA2D3	***DNMT3B***	***GNB1***	***MLL2***	***PCDH15***	*SETBP1*	***ZRSR2***
CALR	*EED*	***GRIN2B***	***MLL3***	***PHF6***	***SF3B1***	
CBFB	***ELF1***	*HNRNPK*	***MLL5***	*PHF8*	***SRSF2***	
CBL	***EP300***	*HRAS*	***MPL***	*PHIP*	*SH2B3*	
CDKN2A	***ETV6***	***IDH1***	*MYB*	*PIK3CA*	*SLC7A8*	
CEBPA	***EZH2***	***IDH2***	*MYC*	***PPM1D***	*SMC1A*	

2.2.3 DNA sequencing libraries and Next generation sequencing

DNA samples underwent library preparation at the Wellcome Trust Sanger Institute, Cambridge, UK. Granulocyte derived DNA samples (n=1033) underwent whole genome amplification prior to library preparation (the TGS1 cohort). Sequencing libraries were generated in a 96-well format, with each sample carrying a unique DNA barcode. Pools of 16 libraries were made and hybridized to RNA baits. Pools of 96 cases were sequenced on two lanes of an Illumina HiSeq2500 machine using 75bp paired-end sequencing. Sequencing data were mapped to the GRCh37/hg19 reference genome using the BWA (Burrows-Wheeler Alignment) algorithm[190]

on default settings as part of the Cancer Genome Project pipeline. Median depth of sequencing was 198x at identified variants – this represents the median coverage across variants after removing PCR duplicates, and so represents the number of unique DNA

molecules sequenced. Table 4 shows values for coverage at important hotspot loci.

Table 4: *Average coverage at hotspot loci*

Gene	Protein	Mean coverage	SD	Median	25% centile	75% centile	Chr	Pos
JAK2	V617F	156	59	153	130	177	9	5073770
DNMT3A	R882	238	131	267	165.5	314	2	25457242
DNMT3A	R882	235	129	263	163.5	309	2	25457243
ASXL1	E635	275	186	338	53	400	20	31022403
MPL	W515	275	163	322	103	381	1	43815008
MPL	W515	274	162	321	102	379.5	1	43815009
U2AF1	Q157	245	151	283	88	340	21	44514777
SRSF2	P95H	255	181	310	32	378	17	74732959
IDH2	R140Q	201	118	230	101	273	15	90631934
SF3B1	K700E	256	93	254	210	297	2	198266834
SF3B1	K666N	204	92	213	138	252	2	198267359
IDH1	R132	182	127	182	151	214	2	209113112
IDH1	R132	186	129	187	155	218	2	209113113

As part of clinical trials or diagnostic work-up elsewhere, samples additionally underwent screening for *CALR* mutations by both exon 9 bi-directional Sanger sequencing (sensitivity of >10%)[52] and exon 9 fragment-size analysis PCR (sensitivity >2%)[191], for *JAK2V617F* by allele-specific PCR[192]

, and *MPL* mutations by PCR (W515 and S505)[193]

2.3 Identification of variants

2.3.1 Single nucleotide substitutions – Caveman

The Cancer Variants Through Expectation Maximisation (CaVEMan) algorithm, developed at CGP and run on the Welcome Trust Sanger Institute farm, was used for detection of single nucleotide substitution variants[194]. As paired constitutional samples were not available, a single designated normal sample was used as comparator. This algorithm uses as its input sorted and indexed binary alignment/mapped (BAM) files with a FASTA reference genome. The algorithm proceeds by (1) splitting the file, (2) iteratively assessing the base pair calls at each genomic locus (maximisation step) (3) merging these analyses to form a complete genome and (4) iteratively assigning a probability to each

possible genotype at each position. If these probabilities are above a threshold they are called as somatic mutations and the output is written to a VCF (variant call format) file.

2.3.2 Small insertions/deletions – Pindel

For insertions and deletions, the pindel algorithm[195] (with in-house modifications) was used, again on the CGP farm using the existing pipeline. The pindel algorithm uses an anchor sequence (i.e. the part of the read mapped to the reference genome) to initiate a search for nearby sequences to which the unmapped portion of the read can be mapped, and in doing so reduces the search space for the unmapped sequence. In-house enhancements of the original algorithm include the exclusion of nonspecific anchors, and the use of average read quality to clip poor quality sequence and improve mapping. These enhancements reduce the false positive rates and enable detection of smaller indels. The output of this algorithm is also written to VCF format.

2.3.3 Filtering of CaVeMan- and Pindel-called variants

CaVEMan and Pindel results were merged into two files, one containing the total set of variants detected, and another including only those that had passed the preset filters carried out by post-processing as part of the CaVEMan/pindel pipeline.

In summary, filters for exclusion by CaVEMan that were called in this dataset included:

<1/3 of alleles with mutant base have base quality >25

<8 mutant reads and these are all located in the last third of the read

Presence of the allele in more than 5% of reads from unmatched normal samples

Mutant alleles fall within a simple or centromeric repeat

Mean mapping quality <21

Mutant allele matches a known dbSNP germ-line variant

>10% of reads covering this position contain an indel

Mutant reads were restricted to one strand direction

Filters for exclusion used for pindel in this dataset included:

Events of >4bp also found in wildetype reads

Events must be in >= 3 in a single strand, or >2 in both

Event must be in >=8% of reads if depth <200, or >=4% if >= 200

Events of <=4bp are discarded in repetitive region (>9 repeats)

WT depth must be >5 and >=8% of total depth.

30637 CaVEMan calls and 128583 pindel calls met these criteria and were carried forward to further post-processing (see Section 2.4).

Reads that contain insertions/deletions may fail mapping due to the resulting discrepancies in the sequence. This may lead to under-estimation of the number of reads carrying the mutation, and therefore of its variant allele frequency. Using in-house algorithms that utilised BLAST (Basic Local Alignment Search Tool) in order to screen unmapped reads for the presence of the mutation, the variant allele fractions of mutations found using pindel were corrected.

2.4 Filtering of variants

Given that matched constitutional samples were not sequenced for the majority of patients, extensive filtering was required to eliminate potential artefactual calls and germ-line SNPs as listed below, in order to enrich for high confidence driver mutations. The filtering methodology used here was adapted from that used in previous studies[189,196]

2.4.1 Removal of germ-line SNPs

Variants were annotated for their presence in the 1000 genome project[197], NHLBI Exome Sequencing Project[198], Wellcome Trust internal 500 exome sequencing dataset, the Ensembl dbSNP database[199], the Exome Aggregation Consortium dataset (MacArthur lab, Broad Institute)[200] and the Catalogue of Somatic Mutations in Cancer (COSMIC) dataset[201–203], and specifically whether they were found to be somatic, the tissues they were found in, and whether any other COSMIC variants were reported affecting the same genomic or amino-acid positions or were within 3 amino-acids of it (the latter criterion was later revised). Variants that matched entries in COSMIC describing somatically acquired mutations at the same genomic or amino-acid position were retained. Some single-nucleotide variants (SNVs) matched a variant in COSMIC by either genomic position or amino-acid position, however, if these variants were not reported as somatic

mutations in COSMIC and had a VAF>40% then they were removed. SNVs prevalent across SNP databases were filtered out. SNVs that were not prevalent in SNP databases, did not match (by exact mutation, genomic location or amino-acid position) to a variant in COSMIC and had a VAF>40% were removed to reduce potential contamination by rare and private germ-line SNPs. SNVs that were not reported in any SNP database, but were within 3 amino-acids of a somatic variant reported in COSMIC were retained if their VAF was less than 40% due to the possibility that these mutations could represent previously unreported somatic events.

2.4.2 Identification of high-confidence somatic calls

Further filtering was carried out as follow:

- Off target calls (including inter-genic or intronic calls) and synonymous mutations were removed.
- CaVEMan variants occurring within regions of germ-line indels and variants located only on one strand (where read depth was >5 on both strands) were also removed.
- Variants <1% allele fraction were removed.
- Variants were further filtered against a panel of 455 samples (blood from patients without known haematological malignancy) that had undergone sequencing on the same platform to remove sequencing artefacts.
- Insertions or deletions that were removed if they were only detected in unidirectional sequencing reads, were located in homopolymer tracts, or were present in sequencing files of a panel of normal samples.
- In parallel to this, known hotspots within myeloid driver genes were screened for low burden canonical mutations. Loss-of-function mutations (as caused by nonsense, frameshift or splice site mutations) in the genes ASXL1, BCOR, CUX1, DNMT3A, TET2, PPM1D, EZH2, GATA2, MBD1, RB1, RUNX1, SH2B3, NFE2, NF1, TP53, CBL, MLL3, and ZRSR2 were also retained provided they were not detected as artefacts of sequencing when compared to the normal panel.

The filtering strategy above was further refined by taking samples that had undergone variant calling using a matched constitutional germ-line sample, and interrogating

differences in variants that would have been called if one were to perform variant calling without the matched constitutional germ-line sample. While this filtering strategy above might still not remove all germ-line variants, the variants included likely reflect those that might be identified from real world diagnostic practice where routine practice involves sequencing of tumour DNA only and not a matched constitutional analysis.

2.4.3 Genotyping of germ-line variants of interest

Using SAMtools mpileup[204] at loci for a pre-specified list of SNPs (chosen for their association with haematological parameters and/or MPN predisposition[130,140,149,159,205–209]), allele fractions for each allele were determined and therefore each patient was allocated a genotype. SNPs which were not adequately covered for the entire cohort were not included in further analysis.

The final list of SNPs is shown in Appendix 1, together with their chromosomal location, phenotypic correlates and references.

2.5 Identification of chromosomal losses/gains/uniparental disomy

Dominant allele fractions were calculated for SNPs across the genome using SAMtools mpileup at loci derived from 1000 genomes project[197] (using Ensembl biomart[210]). B-allele frequency plots for each sample were assessed to determine whether regions of deviation from 50% allele frequency for heterozygous SNPs could be identified. Normalised coverage plots were then assessed in these cases to classify chromosomal changes into gains, losses or copy-neutral loss-of-heterozygosity (CN-LOH) events. These were plots of individual base-pairs for each locus within targeted regions, normalised against the median values for that locus for cohort as a whole (independently for TGS1 and TGS2).

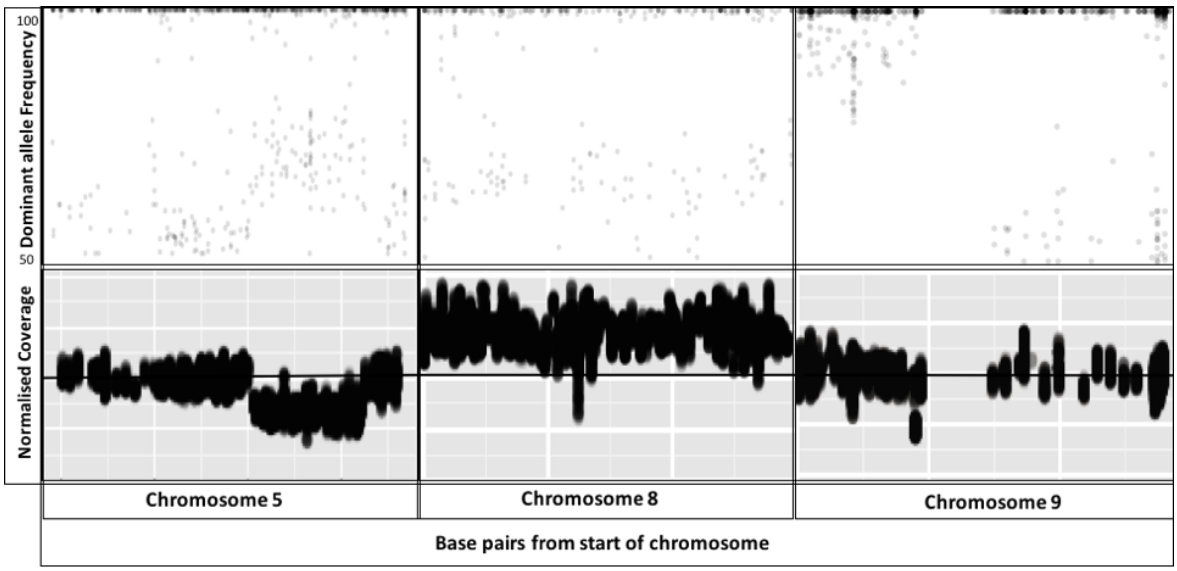


Figure 2: *Detection of chromosomal abnormalities. Dominant allele fractions (top) and normalised coverage (bottom) are show for examples of chromosomal loss on 5q, chromosomal gain (trisomy 8) and CN-LOH of 9p.*

Recurrent abnormalities were detected across chromosomes 4, 5, 7, 8, 9, 11, 12, 13, 14, 17, 18, 19 and 20. Traditional cytogenetic analyses, either from FISH or karyotyping, were available for only 20% of the cohort. In these patients, 88% concordance was present between copy number changes detected by FISH/karyotype and their detection in data from NGS. However, it was not possible to identify chromosomal translocations.

2.6 Single cell derived colonies

5.5mL aliquots of MethoCult H4435 Optimum (StemCell Technologies, Vancouver, Canada) at room temperature were prepared. MNCs were diluted in PBS to a concentration of 6-7x106 cells/ml, and 200µl aliquots of the cell suspension were added to each 5.5ml aliquot of MethoCult. For cytokine experiments recombinant human TGF-a or GRO-a/CXCL1 (R&D Systems, Abingdon, UK) were added to a final concentration of 50, 25 and 10mmol/l, together with an untreated control. Samples were vortexed for 30 seconds. Each MethoCult aliquot was then divided across 3 wells of a 6-well plate at 1.5ml per well, using a 5mL syringe and an 18-gauge needle.

Sterile water/PBS was dispensed between the wells and the and 6-well plates were incubated at 37°C with 5% CO2 and high humidity for 14-15 days. Following in vitro culture, plates were visualized under a dissecting microscope. Single erythroid (BFU-E) or granulocyte-monocyte/macrophage (CFU-GM) colonies were recognized by their morphological characteristics and individual colonies were aspirated and lysed in 50µl of RLT lysis buffer (Qiagen, Maryland, USA) and stored at -80°C.

2.7 Measurement of serum cytokine concentrations

MILLIPLEX MAP Human Cytokine/Chemokine Magnetic Bead Panel Immunology Multiplex Assay (Millipore, HCYTOMAG-60K) was selected to measure levels of 38 cytokines. This method utilises Luminex xMAP® technology, which uses antibody-coated, colour-coded magnetic microspheres each specific to a different analyte. After incubation with the sample these are treated with a biotinylated detector antibody and a Streptavidin-PE conjugate reporter conjugate which is then visualised using a Luminex xMAP flow cytometer. Absolute concentrations are then calculated from MFI using a standard curve for each analyte. The use of specific, independently identifiable beads means that multiple analytes can be measured in a single well.

This work was done together with Dr David Kent, Dr Nina Friesgaard Oebro, Miriam Belmonte and Olivia Harris. The cytokines measured were: Vascular endothelial growth factor (VEGF), soluble CD40 ligand (sCD40L), epidermal growth factor (EGF), Eotaxin, fibroblast growth factor 2 (FGF-2), Flt-3 ligand, Fractalkine, granulocyte-colony stimulating factor (G-CSF), granulocyte macrophage-colony stimulating factor (GM-CSF), growth-regulated oncogene (GRO/CXCL1), interferon alpha 2 and gamma (IFN-α2 and IFN-γ), interleukins (IL) IL1α, IL-1β, IL-1Ra, IL-2, IL-3, IL-4, IL-5, IL-6, IL-7, IL-8, IL-9, IL-10, IL-12(p40), IL-12(p70), IL-13, IL-15 and IL-17, interferon gamma induced protein 10 (IP-10), monocyte chemotactic proteins 1 and 3 (MCP1 and MCP3), macrophage-derived chemokine (MDC /CCL22), macrophage inflammatory protein 1 alpha (MIP1α), MIP1β, transforming growth factor alpha (TGF-α), tumour necrosis factor alpha (TNF-α) and TNF-β.

The kit was run according to manufacturers instructions; briefly, 200µL assay buffer was pipetted into each well of the 96-well plate and left to stand for 10 minutes at room temperature, followed removal of the assay buffer by tapping out. 25µL of assay buffer was added to each well followed by 25µL of standard or control to the appropriate wells. 25µL matrix solution was added to the background, standard, and control wells. Patient serum samples were thawed and brought to room temperature and 25µL of each sample added to appropriate wells. After mixing, 25µL pre-sonicated beads were added, and the plate was incubated overnight at 4°C with agitation. The plate contents were tipped out whilst on a magnet and the plate was washed twice. 25µL detection antibodies were added to each well and the plate was incubated for a further 1 hour at room temperature with shaking. After this, 25µL streptavidin--phycoerythrin was added to each well and the plate

was incubated for another 1 hour at room temperature with shaking. The well contents were then tapped out and the plate washed, and finally 150μL sheath fluid was added. After re-suspension for 5 min, the plate was run on Luminex xMAP (Luminex Corp., Austin, TX). Cytokine concentrations were then determined by xPONENT software (Luminex) on the basis of the fit of a standard curve for MFI versus pg/mL, using values derived from the known reference concentrations supplied by the manufacturer.

2.8 Statistical analysis

The code for most statistical analyses performed is provided in Appendix 2.

2.8.1 Tests of significance and multiple hypothesis testing

Testing for difference in means for continuous variables was performed using student's t test, or Mann-Whitney/Wilcoxon test for non-normally distributed data (t.test() and wilcox.test() in R). Testing for differences in proportions was performed using chi-squared or Fisher's exact test (chisq.test() or fisher.test() in R).

For specific correlative analyses, the Benjamini-Hochberg method was used to adjust p values for multiple hypothesis testing (using p.adjust()). The adjusted values quoted in the manuscript therefore represent q values (that is, the false discovery rate, FDR). This approach was in particular applied to analyses of associations among driver mutations, germ-line variants and MPN subtype at diagnosis.

Where possible however, the tests of association between variables were performed using multivariate analysis to correct for the confounding effects of other variables. Fitting of generalised linear models and logistic regression were performed using glm(), and of linear mixed effects models using lmer() from lme4 v1.1-12.

For the multivariate survival analyses, overfitting was controlled by L2 regularisation through the random effects model (discussed in more detail in Section 5.7). Additionally model validation was performed using leave-one-out cross-validation, and through the use of external datasets.

2.8.2 Bayesian network analysis

The Bayesian network (BN) was constructed by first creating a binary matrix indicating presence/absence of genomic features of interest, with each row in this matrix corresponding to a single patient, and columns to genomic features. A data value of 1

indicates the presence of the feature in the specific patient. The Gobnilp software was then used with default parameters to learn the structure of the network[211]. The parameters included a maximum of 3 parents for each node. The algorithm finds the globally optimal BN within the constraints set, by treating the learning of the BN structure as a linear optimisation problem. This is then solved using integer linear programming via the SCIP Optimization suite 3.2. For the resulting BN we used in-house code written in SWI-Prolog[212] that finds the logic gate combination which maximizes the Fisher exact test p-value when regressing the parents against the child. For a child, its input vector is computed by performing the logic gate operations on its parents' value vectors.

2.8.3 Determination of clone sizes and mutation acquisition order

The proportions of cells carrying each mutation were calculated using the variant allele fractions corrected for any copy number change at the site of the variant, and comparisons were made for each pair of mutations in each patient to determine the order of acquisition as described previously[122,213]. Here, the terms ‘early’ and ‘late’ for describing driver mutations are used to denote relative timing. That is, we cannot know when in chronological time an event occurred; rather we can sometimes infer the relative order in which two events occurred. This then generates data analogous to scenarios that arise in sports statistics, where a league of teams play one another in a series of pairwise encounters, and we are interested in inferring from this series of pairwise events an overall ranking of teams in the league. We adopted methods from sports statistics to generate a relative ranking of driver mutations about whether they occur typically ‘early’ or ‘late’ during disease evolution relative to other driver mutations. R packages BradleyTerry2 (version 1.0-8) was used to generate estimates of relative mutation timing with quasi-variances generated using qvcalc (version 0.9-1).

2.8.4 Bayesian clustering algorithm

We performed clustering on the genomic data alone using a Bayesian approach previously applied to genomic data from patients with AML[214]. In brief, the number of clusters was learned by using a Markov Chain Monte Carlo (MCMC) method to sample from an underlying Dirichlet process, utilising an implementation of the Dirichlet process mixture model (https://github.com/nicolaroberts/hdp, hdp package v0.1.0). The MCMC was run with an initial 500 iterations, which were discarded (“burn-in iterations”), followed by 1000 iterations sampled at intervals of 20. The optimal number of clusters using our dataset of 2035 patients and 54 genomic changes was found to be 7. This number of

categories was unchanged, when the cohort was combined with 536 genomically characterized patients with MDS. From this model, each patient had a probability of being assigned to each of the seven classes, and from these probabilities, a simplified set of rules was generated to allow for patients to be categorised into the group for which they had the initial maximum probability. We included an additional group of patients who were found to have no genomic changes (and who were therefore assigned equal probabilities of belonging to each class).

2.8.5 Survival analyses

The process by which survival models were fitted and integrated is documented in more detail in Chapter 5. Simple Cox proportional hazards models were fitted and Kaplan-Meier analyses performed and plotted using the survival 2.4-1 and rms 5.1.0 packages. Receiver-operator curve sensitivity/specificity analyses were performed using the survivalROC 1.0.3 package. Variable selection was performed using Akaike's information criterion using k=log(n) using the stepAIC function from MASS v7.3-45. Mixed proportional hazards models were fitted using the CoxHD() package (v0.061)[215]

The variables included in the modelling were as follows:

- Demographic: age at diagnosis (continuous variable), gender (male/female)
- Clinical (at diagnosis): history of prior thrombosis (present/absent), splenomegaly (present/absent), haemoglobin concentration (continuous variable), white cell count (continuous variable), platelet count (continuous variable), PV versus ET (for chronic phase patients).
- Genetic: Presence (>1% VAF) /absence of mutations in 33 genes (individual variables) - JAK2 (V617F or exon 12), CALR, MPL (S204/S505/W515 hotpots), ASXL1, BCOR, CBL, CUX1, DNMT3A, EZH2, GATA2, GNAS, GNB1, IDH1, IDH2, KIT, KRAS, MLL3, NFE2, NF1, NRAS, PHF6, PPM1D, PTPN11, RB1, RUNX1, SF3B1, SH2B3, SRSF2, STAG2, TET2, TP53, U2AF1, ZRSR2.
- Cytogenetic: Presence/absence of copy number changes or copy number neutral loss of heterozygosity in 16 regions (individual variables) - Chromosome 1 (1p CN LOH or 1q gain), 4, 5, 7, 8, 9 (9p CN LOH or trisomy 9), 11, 12, 13, 14, 17, 18, 19, 20.

Cohort: Original cohort of patient (e.g. PT-1 cohort, Italian cohort, UK centres)

Individual models were created for transitions from (i) Chronic phase (CP, namely ET, PV and MPNu patients) to death (censored at AML or MF transformation), (ii) MF to death (censored at AML transformation), (iii) CP to MF (censored at death or AML transformation) and (iv) CP or MF to AML (censored at death; transformation from CP to MF was treated as a time-dependent variable). Time 0 was taken to be the time of diagnosis, but for cases of secondary MF where a patient was sampled in CP (data used for modelling the survival fits in (ii) and (iv)), the time of transformation was taken as time 0.

Data was right-censored at the end of follow-up. The median time between diagnosis and sample acquisition was 55 days. In addition, 42% of patients were sampled within a month of diagnosis and 69% of patients were sampled within the first year from diagnosis. This is an important consideration when building prognostic models as a mutation that is present from the time of the sample may not have been present prior to sampling. Therefore, our model left-censors from time-of-diagnosis to time-of-sample for survival analyses. This ensures that the risk associated with any detected mutations is only applied to the period following sampling (since the presence/absence of the mutation is unknown up until that point) and thus, reduces any potential lead-time bias that could be introduced from any patients sampled later in disease.

Proportionality was checked using scaled Schonfeld residuals, both by plotting these against time and testing formally for non-proportionality. For each variable, in each of the individual fits used in the multi-state model, there was no evidence that the proportionality assumption was violated.

The individual transition models are then integrated into a single multistate model, using the principles described previously[215]. Death but also transformation to AML were treated as terminal events. A “leave one out” method was used for internal cross-validation, that is, predictions for each patient were generated using a multi-state model built on data from the remaining patients. External validation was also carried out using a cohort of 417 patients from Italy, for whom a subset of genomic information was available (58% for MF patients and 47% for CP patients). The performance of the model was assessed for each individual transition (CP to death, MF to death, CP to MF, CP to AML and MF to AML) and for overall event-free survival (EFS).

Random effects Cox proportional hazards models were created using the CoxHD package

(v0.0.61) and integration into multistage models using msSurv (v.1.2.2) and rcpp (v0.12.11).

2.8.6 Methods for prognostic model evaluation

Numerous methods for evaluation of prognostic model performance exist, each with their own advantages and disadvantages, and the choice of comparator will also depend on the clinical question being asked. In order to fully characterise a model, and compare it to other possible prognostic models, a number of metrics were calculated, namely: Harrell's c-statistic, Uno's C, the Brier score and absolute prediction error.

Harrell's c-statistic is defined as the probability that a given pair of patients is correctly ranked according to the prognostic model, and therefore evaluates all comparable pairs of patients[216]. A value of 100% represents a fully accurate ranking, while that of 50% would be equivalent to chance. Generally speaking, values of 60-70% are seen in survival analyses.

One limitation of this method is that it only determines whether the model correctly ranks patients by risk, but not the accuracy of the prediction for a given patient. Secondly, if a pair of patients includes a patient right-censored prior to the time of an event in the second patient, these patients cannot be compared, and therefore the c-statistic is subject to the censoring distribution of the individual study. Therefore a modification of the c-statistic, which is independent of censoring, is also included[217]. However, since the comparisons presented in this thesis are of multiple models applied to one single cohort (and one external validation cohort), one would expect this second limitation to be less crucial for the analyses reported here.

The use of the Brier score and absolute prediction error addresses the first limitation of concordance measures mentioned above. The absolute prediction error is defined as the mean difference between predicted and actual outcomes while the Brier score is defined as the mean of the squared difference between the predicted and actual outcome[218]. Therefore, a Brier score of 0 would represent a fully accurate set of predictions, while a score of 0.25 would be equivalent to a coin-toss (in the context of binary outcome). The survival state for censored patients is extrapolated from their last time-point.

From these, briar skill scores can be calculated, defined as the ratio of the briar score to

the “baseline” briar score where the prediction is taken as the baseline frequency of the event in all cases[219]. The accuracy of a prediction depends in part on the frequency of the event (f) being predicted, or more specifically underlying uncertainty (defined as f(1-f) and ranging from 0 to 0.5). Because better (i.e. lower) Brier scores are seen where the uncertainty is lower it is not possible to directly compare scores across outcomes or time-points with different levels of uncertainty. For this reason, the uncertainty for each prediction was also calculated for reference.

2.8.7 Decision tree/Recursive partitioning and Random forest classifications

Creation of alternative classifications for overall or progression-free survival were performed using decision tree algorithms implemented in party v.1.1-2 and rpart v.4.1.10. Random forest analyses were performed using randomForest (v.4.6-12) for classification, and randomForestSRC v.2.4.1 for survival analyses. These algorithms were also used to identify candidate cytokines that were best able to categorise patients by MPN subtype.

2.8.8 Other analyses and statistical packages

The following functions and packages in R (v.3.2.2) were additionally used for data processing and visualisation: lsr v0.5, data.table v1.1, DT v0.2, tidyr v0.6.1, Hmisc v4.0-2, ggplot2 v2.2.1, hilbertVis v1.26, lattice v0.20-34. The arules package (v. 1.5-0) was used for frequent pattern mining.

3. Genomics of Myeloproliferative Neoplasms

3.1 Introduction

Myeloproliferative neoplasms are generally considered genomically simple, with many fewer detected mutations per patient than solid malignancies and most other haematological malignancies. Furthermore the set of genes mutated is also generally small, in that most cases have mutations in one of the phenotypic driver mutations, JAK2, CALR or MPL, without mutations of other affected genes[52,66,95].

Despite this, however, the associations between the additional mutations and chromosomal abnormalities, and between these genomic changes and phenotype and outcome are not well characterised. This is because most genomic studies in MPNs have either focussed on myelofibrosis alone, have small sample sizes, have been restricted to a small number of genes and/or have focussed on mutations alone without assessing the presence of chromosomal changes. An understanding of these questions would inform our knowledge of how genomic changes interact and give rise to particular phenotypes, inform our ability to accurately define and classify disease entities and improve our ability to prognosticate and therefore make management decisions.

Table 5: Number of patients with ET, PV, MF or other MPN diagnoses sequenced, with associated demographic and clinical parameters, and frequencies of *JAK2*, *CALR* and *MPL* mutations, or absence of *JAK2*, *CALR* or *MPL* ("triple negative"). MPNu=MPN undefined.

Diagnosis	**ET**	**PV**	**MF**	**MPNu**	**Other***
Number of patients	1321	356	309	14	35
Median age (years)	54	62	63	54.5	57
Female (%)	59.1	45.2	35.6	50	34.3
Median Hemoglobin concentration (g/l)	140	179	116	123	128
Median white cell count (x10^9/l)	9.1	10.6	9.4	7.35	7.95
Median Platelet count (x10^9/l)	862	480	329	356	218
***JAK2*^{V617F} mutated (%)**	54.7	99.7	70.2	64.3	17.1
***JAK2* exon 12 mutated (%)**	0	0.02	0	0	0
***CALR* (52bp deletion/Type 1-like) mutated (%)**	16.2	0	13.6	21.4	0
***CALR* (5bp insertion/Type 2-like) mutated (%)**	10.8	0	4.9	7.1	0
***MPL* mutated (%)**	4.5	0	4.5	0	0
Triple negative (*JAK2/CALR/MPL* unmutated)	15	0	8.1	14.3	82.9

* Comprised systemic mastocycosis (n=7), refractory anemia with ringed sideroblasts (n=2), post-PV acute myeloid leukemia (n=1), myelodysplasia/MPN/SM overlap (n=1), idiopathic erythrocytosis (n=11), Hypereosinophilic syndrome (n=2), chronic myelomonocytic leukemia

(n=6), chronic eosinophilic leukemia (n=1), atypical chronic myeloid leukemia (n=1), and 2 patients with unspecified MPN diagnoses.

We therefore utilised a large cohort, with representation of patients with both chronic phase (ET and PV) and myelofibrosis and used a relatively unbiased targeting approach that included not only genes known to be mutated in myeloid malignancies, but also a number of putative drivers identified in previous studies. The cohort of 2035 patients comprised 1326, 355 and 310 patients with ET, PV and MF respectively and 50 patients with other MPN diagnoses. Clinical characteristics and diagnoses are detailed in Table 5.

3.2 Mutations and chromosomal aberrations in MPNs

3.2.1 High confidence mutation and chromosomal aberration calls

As discussed above (Section 2.4), a number of criteria were used to define a variant called by Caveman (for substitutions) or Pindel (for insertions/deletions) as a probable somatic driver mutation. As the somatic nature of these mutations could not be determined definitively, as a matched constitutional sample was not available (and genotyping of single cell colony derived colonies, done in some cases, was not feasible) for the vast majority of cases, filtering against publically available SNP databases was used to exclude known germ-line variants.

The COSMIC database was also used to identify those mutations that had previously been shown to be somatic, and the recurrent nature of these mutations was used as an indicator that they acted as oncogenic drivers (rather than functionally inert passengers), in the absence of a functional assay. For genes known to be affected by protein truncating or loss of function mutations (e.g. TET2, ASXL1, EZH2, NFE2), any nonsense/frameshift/splice-site mutation was included regardless of COSMIC matches. Initially, missense mutations were required to be within 3 amino acids of a recurrent COSMIC match, previously shown to be somatic. However, subsequent examination of the cohort that had undergone whole exome sequencing, for whom matched constitutional samples were available revealed a high proportion of germ-line variants even using this criterion. Therefore, for missense mutations, only those at the same amino acid site of a COSMIC match, previously demonstrated to be somatic, were included. One caveat in the discussion that follows therefore is that this dataset may still include rare germ-line events because of these limitations.

Only those mutations that were mutated in at least 5 patients in our cohort (0.25%) were

taken forward for further analysis, as genes mutated below this threshold were unlikely to play a significant role in MPN pathogenesis and would be insufficiently powered to draw conclusions about genotypic or phenotypic correlations. The variants detected are found in Appendix 3. We found 33 genes to be mutated in at least 5 patients. Only JAK2, CALR, TET2, ASXL1 and DNMT3A were mutated in >5% of patients (Figure 1A). As expected, JAK2V617F and MPLW515 mutations represented the vast majority of JAK2 and MPL mutations, while all CALR mutations were +1 base pair frame shift mutations in exon 9.

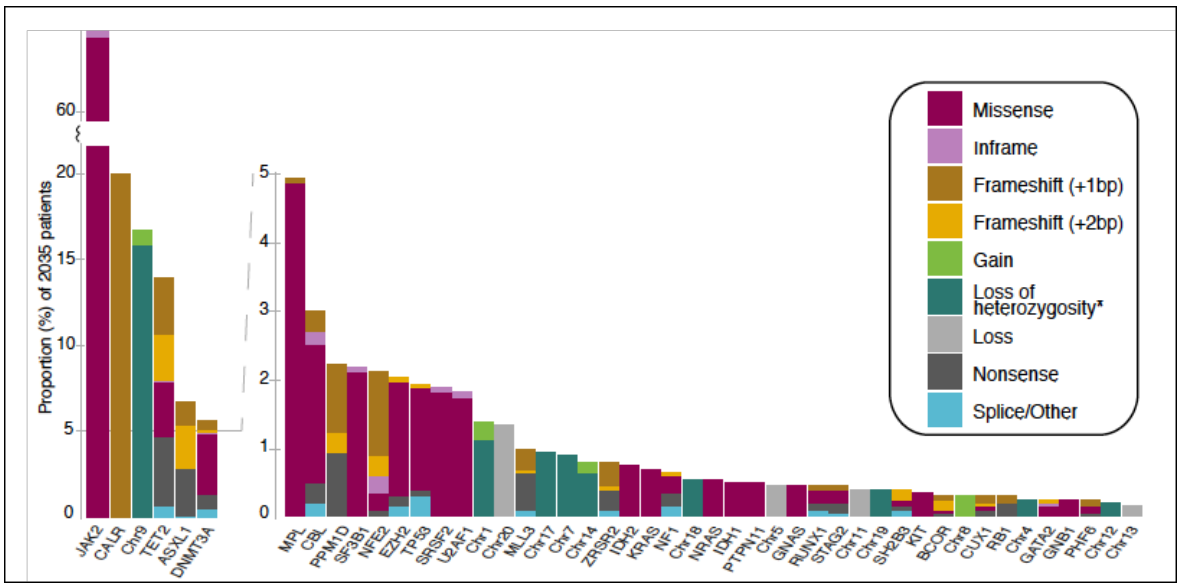


Figure 3: *Frequencies of recurrent somatic mutations and chromosomal changes. *In some cases it was not possible to confidently distinguish copy number neutral loss of heterozygosity from chromosomal loss, and therefore the the term loss of heterozygosity is used to denote both.*

Other mutated genes included those involved in cell surface receptor signalling (MPL, KIT, SH2B3, CBL, GNAS, GNB1), epigenetic regulators (EZH2, IDH1, IDH2, MBD1, MLL3, PHF6, STAG2), spliceosomal components (SRSF2, SF3B1, U2AF1, ZRSR2), members and regulators of the RAS pathway (NRAS, KRAS, PTPN11, NF1), transcription factors (NFE2, GATA2, RUNX1) and regulators of cell cycling and apoptosis (TP53, PPM1D, RB1, BCOR).

The commonest chromosomal changes in MPN patients were (i) uniparental disomy (UPD) of chromosome 9p (16.1% of patients), (ii) gain or UPD of chromosome 1 (1.4% of patients) and (iii) loss of 20q (1.4% of patients) (Figure 3).

3.2.2 Mutation burden

Including phenotypic driver mutations, there were a mean of 1.3, 1.5 and 2.1 mutations

detected in patients with ET, PV and MF respectively, a trend consistent with previous observations. Chromosomal aberrations were detected in 8, 55 and 45% of patients respectively.

Heterozygous JAK2V617F occurred as the only detectable driver mutation in 58%, 33% and 14% of patients with JAK2-mutated ET, PV and MF. Similarly, heterozygous CALR or MPL mutations were found in isolation in 56% and 30% of CALR/MPL-mutated ET and MF patients respectively. Excluding loss of heterozygosity (LOH) for JAK2V617F, 33.6, 56.1 and 20.3% of JAK2-mutated ET, PV and MF patients respectively had no additional somatic drivers. LOH at CALR/MPL loci (chromosomes 19p and 1p respectively) was comparatively infrequent.

Overall therefore, mutation burdens across ET, PV and MF are relatively low and, in a large number of cases, heterozygous or homozygous mutations of a phenotypic driver gene do not appear to require cooperating (coding) mutations to give rise to a disease phenotype, even in cases of accelerated phase disease (MF).

3.2.3 Hotspot mutations and comparisons to other malignancies

As noted above the vast majority of JAK2 mutations affected the V617 hotspot (V617I in one case and V617F in the remainder) and those of MPL largely affected the 515 hotspot. MPL W515 mutations were most commonly W515L or W515K (68 and 17% of cases respectively) but substitutions to W515R, A, G and S are were also seen. Non-canonical mutations of JAK2 and MPL are discussed in section 3.2.4.

NRAS, IDH1 and 2, SRSF2, SF3B1 and GNB1 were also exclusively or highly enriched for hotspot mutations with those affecting IDH1, IDH2, SRSF2 and GNB1 restricted to single hotspots previously reported in myeloid malignancies (R132, R140, P95 and K57 respectively). Additionally DNMT3A R882H/C mutations constituted 26% of the mutations of this gene.

As has previously been reported in AML and MDS, as well as malignant melanoma, NRAS mutations predominantly affect either G12/13 or Q61. Data from the COSMIC database[203] (restricted to these hotspot regions) reveal that G12/13 mutations predominate in AML, MDS, JMML and MPN/MDS-overlap cases (73-93% of samples, n=1586). Interestingly, the converse is true in malignant melanoma (n=1838) and plasma cell myeloma (n=184) where Q61 hotspot mutations predominate (88 and 78% respectively).

In contrast, no cases of Q61 mutations were seen in this cohort, although two cases had mutations affecting A59 and G60 (mutations affecting these amino acids were observed in approximately 6% of MPN/MDS-overlap cases from the COSMIC database but were seen in <1% in the other conditions assessed) while all remaining cases had G12D mutations.

SF3B1 mutations were also almost entirely restricted to two hotspots – K700 (K700E in 13 cases) and K666 (K666N/Q/R/T in 20 cases). Data from the COSMIC database reveal that cases with MDS (whether with ringed sideroblasts or not, n=560) and CLL (n=180) predominantly carry mutations at the K700 rather than the K666 hotspot (approximately in a ratio of 80:20), whereas the ratios seen in AML (52 vs. 56 cases) and in this cohort are skewed more towards the K666 hotspot.

These finding of mutations at known hotspots in these genes serves to validate the sequencing and variant filtering methodology used here, but also suggests that mutation at different hotspots of NRAS and SF3B1 may have different phenotypic consequences across myeloid and other malignancies.

3.2.4 Non-canonical JAK2 and MPL mutations

Complete coding sequence analysis of JAK2 and c-MPL identified non-canonical variants in 67 cases, affecting 18 amino acids in JAK2 and 12 in MPL, of which 6 and 4 sites respectively had recurrent mutations. Although the somatic and oncogenic nature of these variants was not necessarily clear for a number of these cases, a number of variants appear to be of potential interest (Figure 4):

(i) JAK2 R683G and E627A were identified as the sole detectable mutation in two ET patients, one of whom presented in childhood. Both have been reported in acute lymphoblastic leukemia (ALL) where they result in JAK2 constitutive activation.

(ii) JAK2R867Q/W was identified in 2 ET patients, again in the absence of any other phenotypic driver mutation, and has previously been associated with familial thrombocytosis[23]. Variant allele frequencies did not allow differentiation between germ-line or somatic acquisition in these cases.

(iii) MPL S505N and MPL S204P were identified in 5 and 6 ET patients respectively and have been reported previously as somatic phenotypic driver events in ET[48-50]. In 9 of these 11 cases, there were no other phenotypic driver mutations. Additionally, MPL S204P co-occurred with UPD affecting chromosome 1p in two

cases, further suggesting an oncogenic role for this variant.

(iv) JAK2 N1108S (found in 12 cases), R340Q (5), R564L (3), MPLY591D (3) and R592Q (3) have been reported in JAK2/MPL-unmutated MPNs[49]. However, in this cohort these variants more commonly occurred alongside canonical phenotypic driver mutations.

(v) JAK2 I951T (found in 6 cases) has not previously been reported as a somatic or germ-line variant. 5/6 cases carried JAK2 V617F mutations, and 4/6 cases had a PV phenotype. As this variant was restricted to the Florence cohort, this raised the possibility that I951T was a rare SNP, potentially enriched in Tuscany, and further analysis was carried out by Dr Guglielmelli on these patients, suggesting this was indeed a germ-line variant.

Because of the unclear somatic and/or oncogenic nature of a number of these non-canonical mutations, only patients with MPL W515/S505/S204 or JAK2V617F/Exon12 mutations were classified as MPL- and JAK2-mutated respectively for subsequent analysis.

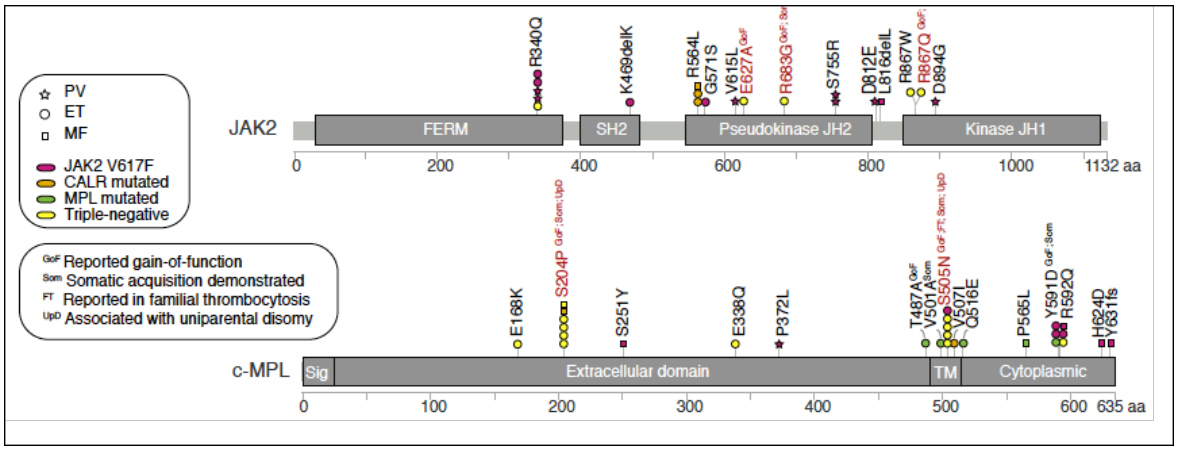


Figure 4: *Locations of non-canonical mutations of non-canonical variants found in JAK2 and MPL. Marker shapes denote the phenotype of the patient the mutation was found in, and colours denote the presence of additional JAK2, CALR or MPL mutations.*

3.2.5 Novel mutations in PPM1D and MLL3

Protein phosphatase Mg2+/Mn2+ dependent 1δ (PPM1D), also known as wild-type p53 induced phosphatase (WIP1), is a serine/threonine protease that acts as a regulator of the DNA damage response. Frame-shift and nonsense mutations in the terminal exon of PPM1D have been detected in tumours of the central nervous system, as well as in blood from both patients with breast/ovarian tumors[220,221] and healthy ageing individuals with

clonal hematopoiesis[222,223]. Some reports have suggested that these mutations may be germ-line. Protein truncating exon 6 mutations have been found to confer a longer half life to the protein, pheno-copying TP53 mutation, with lack of cell-cycle arrest in response to DNA damage.

We identified PPM1D mutations in 38 patients (1.9%) with a similar distribution to those reported in solid malignancies (Figure 5), and where possible these were confirmed on Sanger sequencing. Genotyping of single-cell derived colonies (Figure 5A), as well as mutation timing analysis (discussed in 3.4.1) demonstrates that PPM1D can occur secondary to JAK2.

In 10 patients, *PPM1D* mutations were identified in a later sample post trial entry and were not present at the earlier time point. All these patients had intervening exposure to hydroxycarbamide. Of 17 patients where *PPM1D* mutations were detectable in the earliest sample available, the time between diagnosis and sampling was significantly longer than in patients without *PPM1D* mutations. 12 of the 17 patients had received hydroxycarbamide before the earliest sample was obtained. However, *PPM1D* mutations were also detected at, or within a month of, diagnosis in 20 cases, and in one case in a sample that was taken 7 years after diagnosis with no hydroxycarbamide exposure.

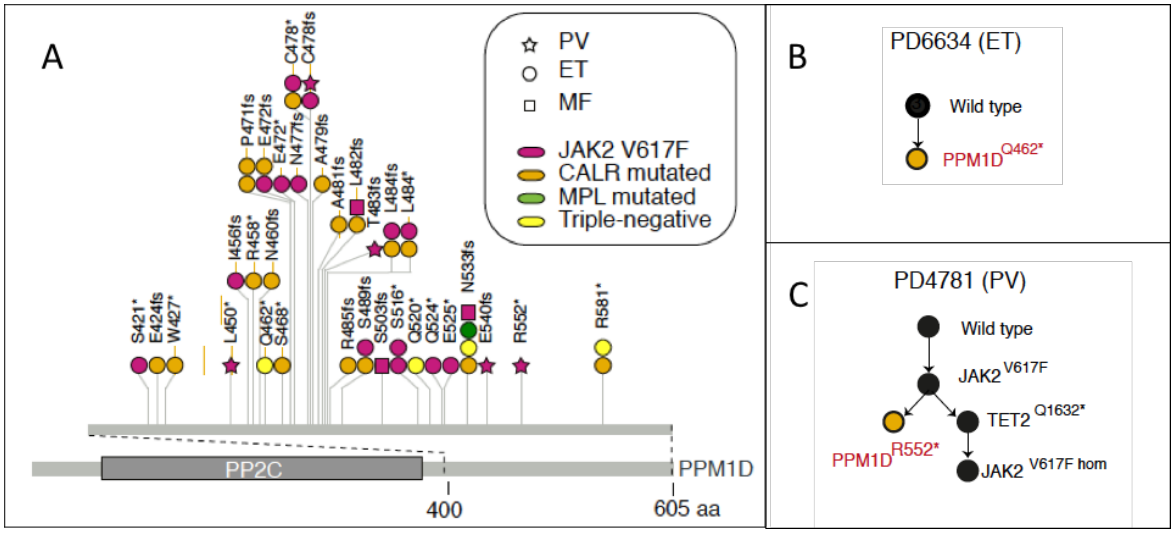


Figure 5: *Mutations detected in PPM1D. Figure A shows their restriction to the terminal exon. Marker shapes denote the phenotype of the patient the mutation was found in, and colours denote the presence of additional JAK2, CALR or MPL mutations. B and C show examples of clonal hierarchies, where PPM1D mutations were found as the sole mutation, together with a wild type clone, or were found subclonal to JAK2 mutations.*

Mutations in MLL3 (Figures 1A and Table S3) were detected in 20 patients (1.0%, of

whom six had 'triple-negative' MPN), and were exclusively nonsense or frame-shift mutations, as have been reported in AML[224]. Unlike PPM1D mutations, these were distributed throughout the length of the gene

3.2.6 "Triple negative" patients

Triple negative (TN) ET and MF are entities of particular interest, since it is often unclear in these cases what is driving the disease phenotype. Therefore genomic analysis has the potential to reveal novel phenotypic drivers in these patients. Conversely, comprehensive genomic analysis may fail to demonstrate a clonal marker, raising the possibility that, for that patient, the myeloproliferative phenotype is not driven by a clonal process. Instead myeloproliferation still may be cell-intrinsic, e.g. due to germ-line variation (as in the familial thrombocytoses/erythrocytoses or high affinity haemoglobin variants) or cell-extrinsic, e.g. due to inflammation, Tpo or Epo hypersecretion, or other micro-environmental changes.

The majority (82%) of triple negative cases of ET had no detectable coding mutations in the set of 33 recurrently mutated genes, and most (67%) of the remaining patients were found to have only one mutation. 3 had non-canonical mutations of MPL (in one case associated with Chromosome 1p UPD), and 6 had non-canonical mutations of JAK2. MLL3 was the most frequently mutated gene of those with a clonal marker (n=6). PPM1D (n=4), TET2 (n=3) DNMT3A (n=3), and 20q- (n=2) were also seen. Of these however, mutations of MLL3 are over-represented in TN ET cases (OR 3.0, 95% CI 0.95-8.5, p-value 0.03), in contrast to TET2 and DNMT3A where they are under-represented in TN disease relative to JAK2/CALR/MPL-mutated ET cases (p-values <0.0001 and 0.006 respectively). This raises the possibility that MLL3 mutations may themselves be sufficient to drive a clonal proliferation, rather than simply being bystander markers of clonal expansion due to another cause.

Conversely, TN MF cases had an average of 1.5 mutations per patient (although 42% still had no detectable mutations), and the majority of those with a clonal marker had multiple mutations, most commonly ASXL1, NRAS, CBL, EZH2 and TET2. NRAS, in fact, appeared to be over-represented in TN MF cases compared to JAK2/CALR/MPL-mutated cases (OR 8.7, 95% CI: 1.69-40.4, p-value=0.005). Isolated SF3B1 was found in 2 cases.

3.2.7 Putative drivers from targeted genome sequencing.

31 additional genes were sequenced as part of the overlap of the TGS1 and TGS2 RNA

bait sets and were found to be recurrently mutated. These data were further interrogated in order to determine whether any putative drivers could be identified, retaining only missense/inframe mutations if present, and shown to be somatic, at the same site in the COSMIC database. A summary of a number of recurrently mutated genes is shown in Figure 6. By definition, the evidence for recurrent oncogenic mutations in these genes was more limited compared to the 33 analysed above, and there was a predominance of splice site mutations of uncertain significance. With the exception of ATRX, PTEN, EP300, BRAF, MLL5, DNMT3B, NOTCH2 and MLL, the VAFs of the variants found in these genes were predominantly >40%, suggesting a large number of private SNPs are still contained within this dataset.

ATRX mutations, seen in 19 patients here, have been reported in patients with MDS who have acquired alpha-thalassaemia/HbH disease, however none of the variants here overlap with those seen in MDS[225]. EP300 and CREBBP mutations are also reported in haematological malignancy[226], but again no overlap is seen with previously reported mutations, and the significance of these mutations is unclear. Overall, none of the genes shown in Figure 6 showed a convincing pattern of mutation or significant overlap with other myeloid malignancies. However this set may still include genes that it would be worth screening for in future sequencing studies.

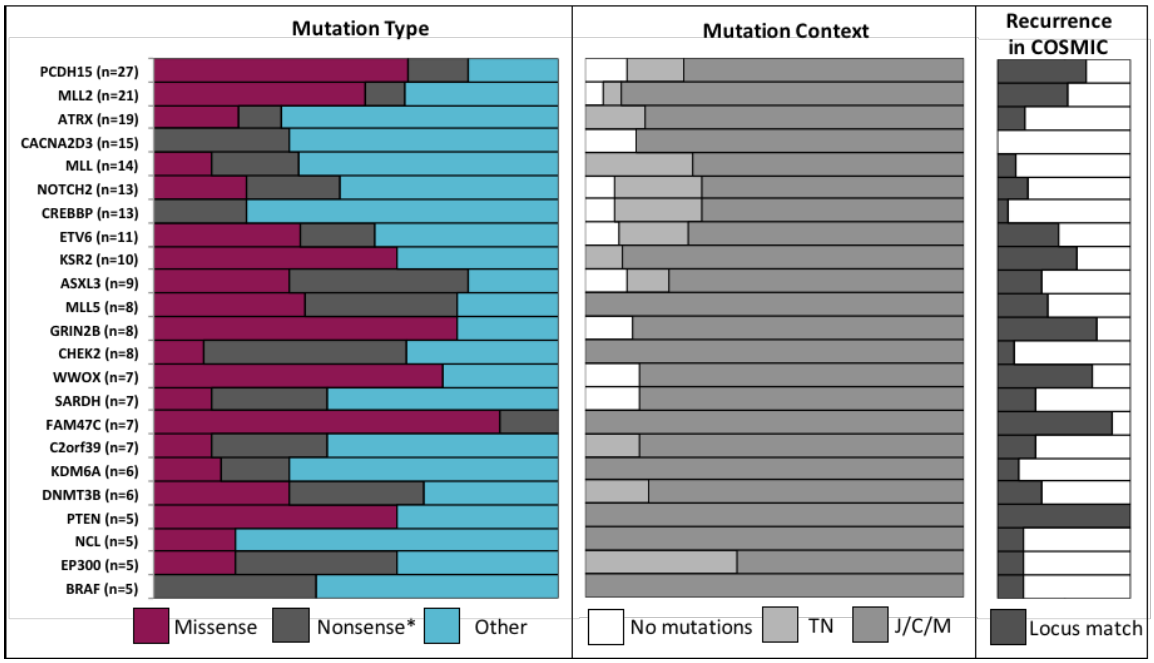


Figure 6: *Candidate driver genes not included in analysis, showing the frequency of each, the proportion of missense, nonsense (*including frameshift mutations) and splice site mutations detected, whether they co-occurred with JAK2, CALR or MPL (J/C/M) or other mutations ("No mutations" signifies the mutation in this gene was the sole mutation*

identified for that patient, whereas "TN" denotes triple-negativity but with other detected mutations), and the number of variants that had matches in the COSMIC database at the same amino acid or base pair.

3.2.8 Putative drivers from whole genome sequencing.

Samples from at least two time points, together with constitutional samples for 35 patients were submitted for whole genome sequencing by Dr Jyoti Nangalia and analysed using pipelines at the Wellcome Trust Sanger Institute.

Initial analysis of this data was carried out as part of this study, but further analysis by Dr Nangalia is ongoing. Firstly, 4 patients were excluded from analysis as pile-up of hotspot loci demonstrated evidence of contamination of constitutional controls. Then, the VAFs of each mutation were compared between each sample and the respective constitutional control using Fisher's exact testing and only those that were significantly higher were kept for further analysis. 39,990 mutations were detected that met this criterion, 15,577 corresponding to a set of 3444 genes. Genes were then ranked according to number of mutations, the recurrence of specific variants, number of mutations relative to gene size, and relative VAF by patient. JAK2 and IDH2 mutations ranked highly according to these criteria.

Other ranking genes included *IKZF2*, *NKD2, ZNF733P* and *ARSD*. A recurrent non-coding mutation in *IKZF2* was observed in two patients with VAFs of 6-28%. *IKZF2* encodes Helios, a member of the Ikaros zinc-finger protein family. While *IKZF1* (Ikaros) mutations are reported in myeloid and lymphoid malignancies[118], Helios mutations have not been described. *ZNF733P* is a zinc-finger protein pseudogene, mutations of which are not previously reported in malignancy. Nkd2 is a regulator of the WNT-pathway that regulates TGF-alpha secretion (in epithelial cells) and has previously been shown to be dys-regulated in a number of malignancies, including AML[227,228]. *NKD2* was mutated in 4 patients with VAFs of 14-46%.

3.3 Patterns of co-mutation

3.3.1 Frequent pattern mining

Identifying recurrent combinations of drivers is of potential interest as it may allow us to pinpoint cooperating events and potentially identify genetically-defined entities. In sections 3.3.1-3.3.3 we discuss sequential refinement of the assessment of patterns of co-

mutations.

Considerable patient-to-patient variability in driver mutations was evident, with 336 different combinations of driver mutations and/or chromosomal events observed, of which only 33 combinations were recurrent in at least 5 cases. Most commonly JAK2 mutations were an isolated event (n=614), CALR mutations were an isolated event (n=258), no mutations were detected (n=192), or JAK2 mutations were found in combination with 9p UPD or trisomy 9 (n=168). These four simple findings made up 61% of the examined cases. Heterozygous JAK2 mutations in combination with TET2 or DNMT3A mutations were the next most common findings (74 and 40 cases respectively).

However, crude enumeration of the combinations of genomic events seen misses recurrent patterns. For example, JAK2+ASXL1+EZH2 (n=2), is counted separately to ASXL1+EZH2+$7q^-$ (n=1), when in fact the combination of ASXL1+EZH2 occurs 16 times in the cohort (and is over-represented beyond the frequency expected by chance). For this reason, frequent pattern finding algorithms were used (implemented in the R package *arules*). The most frequent combinations of 2, 3 and 4 events are shown in Table 6.

This analysis provides a set of "rules" connecting the left-hand side (LHS) and right-hand side (RHS) expressions and a number of metrics describing them. These include (i) support (the proportion of all cases where the LHS gene combination occurs with the RHS expression), (ii) confidence (the proportion of times the LHS occurs with the RHS; namely the positive predictive value) and (iii) lift (the ratio of the observed support against the support expected by chance, i.e. if LHS and RHS are independent events). For example, in 3/5 cases of TP53 mutation and 5q-, 17p- is also seen. Namely the support is 0.0015 (3 of the 2035) total cases and confidence 0.6, an association that occurs significantly more often than expected by chance (lift 64.3).

Table 6: *Most frequent and highest confidence combinations of 2, 3 or 4 co-mutations, with support, confidence and lift for associative rules between left-hand side (LHS) and right-hand side (RHS) expressions.*

LHS	RHS	Support (n)	Confidence	Lift
9pUPD	JAK2	0.169 (n=344)	0.99	1.54
TET2	JAK2	0.094 (n=191)	0.76	1.19
ASXL1	JAK2	0.045 (n=91)	0.71	1.10
DNMT3A	JAK2	0.041 (n=83)	0.73	1.14
TET2	9pUPD	0.033 (n=68)	0.27	1.6
TET2	CALR	0.021 (n=42)	0.17	0.82
ASXL1	TET2	0.017 (n=34)	0.26	2.15
NFE2	JAK2	0.015 (n=31)	0.76	1.17
SF3B1	JAK2	0.015 (n=30)	0.70	1.08
JAK2, TET2	9pUPD	0.033 (n=67)	0.35	2.06
JAK2, ASXL1	TET2	0.012 (n=25)	0.27	2.24
JAK2, NFE2	9pUPD	0.011 (n=22)	0.71	4.17
ASXL1, 9pUPD	JAK2	0.011 (n=22)	1.00	1.55
DNMT3A, 9pUPD	JAK2	0.008 (n=17)	1.00	1.55
JAK2, DNMT3A	TET2	0.007 (n=14)	0.17	1.37
JAK2, EZH2	9pUPD	0.005 (n=11)	0.44	2.59
JAK2, ASXL1	U2AF1	0.005 (n=10)	0.11	6.04
JAK2, ASXL1	SRSF2	0.005 (n=10)	0.11	5.88
JAK2, SRSF2	TET2	0.005 (n=10)	0.37	3.01
JAK2, EZH2	ASXL1	0.005 (n=10)	0.40	6.31
JAK2, TET2, SRSF2	ASXL1	0.003 (n=6)	0.60	9.47
JAK2, EZH2, 9pUPD	ASXL1	0.003 (n=6)	0.55	8.60

There are only 10 rules that have a support of at least 10 patients and confidence of >0.8. 6 of these indicated the strong association of 9pUPD/trisomy 9 (alone or in combination with other mutations) with JAK2 mutations. A strong association between 1p UPD and MPL mutations is also highlighted, with a lift >20. The only other common (n>=10, confidence>50%) associations that do not include co-mutation with JAK2 are those between NFE2 mutations and 9pUPD/trisomy 9 (23 cases, confidence 0.56, lift 3.3), 20q- and CALR (17 cases, confidence 0.61, lift 3.0), and 17pUPD and TP53 mutations (11 cases, confidence 0.58, lift 31.8).

The frequencies of individual mutations and combinations of two mutations is available in an interactive format at https://cancer.sanger.ac.uk/mpn-multistage together with an online implementation of the prognostic modelling developed as part of this study (see Section 4.4.10)

3.3.2 Odds ratios, significance-based and Bayesian network analysis

In order to quantify the strength of associations and their statistical significance, as well as account for negative associations, odds ratios and chi-squared/Fisher's exact testing was used. Given there are 50 mutation and chromosomal events examined, there are 50x50/2=1250 pairwise comparisons. Correction for multiple hypothesis testing was therefore performed, using Benjamini-Hochberg's method to generate values for the false discovery rate (FDR). This identified 47 associations between genomic events with a FDR of <0.05, 8 of which were negative associations. These are shown in Figure 7A.

Independently, Bayesian network analysis was performed from this dataset by Nicos Angelopolous. This method allows the creation of a directed, acyclic graph representing the conditional dependencies between the 50 genomic variables (namely how the presence of a given mutation, or combination of mutations affects the probability of another mutation being present). This complemented the more basic analysis described above in that it allows "pruning" of the network. For example, JAK2 mutations were negatively correlated with 19pUPD, but this is likely to be solely due to the strong negative correlation between JAK2 and CALR mutations, and the strong association between CALR mutations and 19pUPD rather an independent association in its own right. Figure 7B shows a network pruned utilising the dependencies derived from Bayesian network analysis.

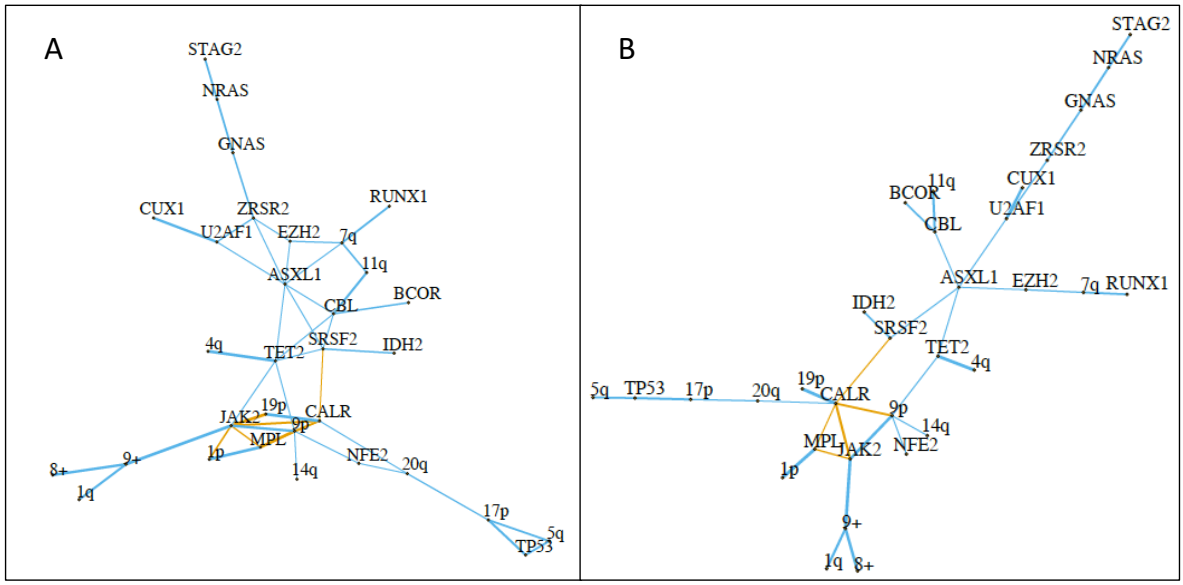


Figure 7: *Networks of co-mutation or mutual exclusivity (A) Naïve model showing all significant associations found in pairwise comparisons, (B) Simplified network, encorporating the results of Bayesian network analysis. Co-mutation above frequency expected by chance in blue, below expected frequency in orange.*

This analysis identified several patterns:

(i) JAK2, CALR and MPL mutations were mutually exclusive (with only rare cases of co-occurrence), confirming the functional redundancy in their pathological mechanisms.

(ii) Multiple genes were strongly associated with LOH at their respective loci, including JAK2, CALR, MPL, TET2, EZH2 and CBL, suggesting that their

increased mutant allele burden conferred a selective advantage.

(iii) Both truncating mutations in NFE2 and TET2 mutations significantly co-occurred with LOH for JAK2V617F (9pUPD and trisomy 9), RUNX1 with 7q-, and CALR mutations with 20q, suggesting that chromosomal deletions/UPD can confer an advantage in combination with other mutations via mechanisms other than increased allele burden/LOH.

(iv) ASXL1 mutations were commonly co-mutated with a variety of additional driver mutations in CBL, EZH2, SRSF2, U2AF1 and NRAS. Within the group of ASXL1-mutated patients, there was mutual exclusivity for SRSF2, EZH2 and U2AF1.

(v) TP53 mutations co-occurred not only with 17p UPD/del but also 5q-. (17p with 5q- occurs in all 3 cases with TP53 with lift 55 and p-value <0.0001, and 5q- occurs with TP53 mutations in 5/9 cases with lift 30.5, p-value <0.0001)

(vi) While the co-occurrence of TET2 with ASXL1 mutations (OR 2.8) and with 9pUPD (OR 2.1) were seen significantly more frequently than expected by chance (p-value<0.0001), the finding of ASXL1 mutation with 9pUPD was rare in TET2-mutated patients (OR 0.24, p-value=0.01).

3.3.3 Bayesian Dirichlet process based approach

Although a number of patterns of clustering have been demonstrated in the above analysis, it remains unclear whether distinct genomic entities/classes exist within this data, into which patients can be classified, and if so how many such classes exist.

To address this question, a Bayesian clustering algorithm, previously utilised in an AML cohort[214], was used to identify possible genomically defined clusters. Since the number and size of any underlying genomic classes is not known *a priori*, potential solutions for the underlying distributions are "learned" by iterative random sampling (using a Monte Carlo Markov Chain method) from an underlying distribution of possible probability distributions (a Dirichlet distribution). The end result is a set of probabilities for the presence/absence of each mutation defined for each class, from which the probability for each patient belonging to each class (i.e. that their individual combination of mutations arising from each classes' probability distribution) can be derived.

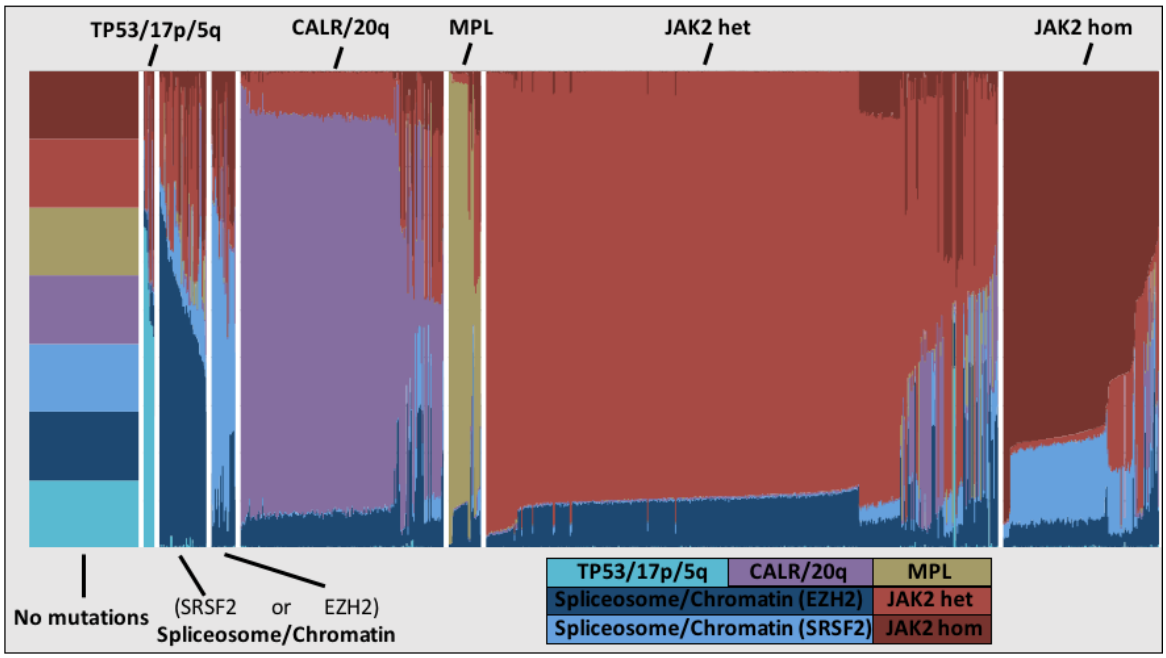


Figure 8: *Individual patient probabilities for allocation to each of the 7 possible classes. Each vertical coloured line represents each of the 2035 patients as is coloured according to the probability of allocation to each class. Patients with no mutations (left) are shown as having an equal probability of being assigned to each class.*

This method consistently identified 6-7 classes. Unsurprising, patients with CALR and MPL mutations separate out into two classes, distinct from those with JAK2 mutations, and JAK2-mutated patients can be separated into those with homozygosity or not. Further classes were predominantly defined by either the presence of TP53 mutations, or two classes by the presence of mutations in spliceosomal or chromatin regulator genes, but in most cases additionally had mutations of JAK2, CALR or MPL. Finally, there remain a subset of patients with no detectable genomic changes ("myeloproliferation with no known driver mutation"), and those with changes not specific to a single class ("myeloproliferation with other driver mutation"). Figure 8 shows the probabilities allocated to each patient for belong in each of the 7 classes and Figure 9 shows the groups of mutations most specific for each group, together with their frequencies and associated odds ratios for co-mutation.

As is evident from figures 8 and 9, classes are not entirely mutually exclusive and an individual patient's allocation was not always clear-cut. This contrasts with a similar analysis carried out for patients with AML, where there were greater numbers of mutually exclusive lesions. In order to allow patient classification, we therefore devised a simple hierarchical classification system which prioritised the TP53-aneuploidy and splicing-chromatin defined categories (which were less common, defined by specific sets of rarer

genomic events and potentially clinically relevant but were more likely to overlap with other categories) over the broader, less overlapping JAK2/CALR/MPL-defined groups.

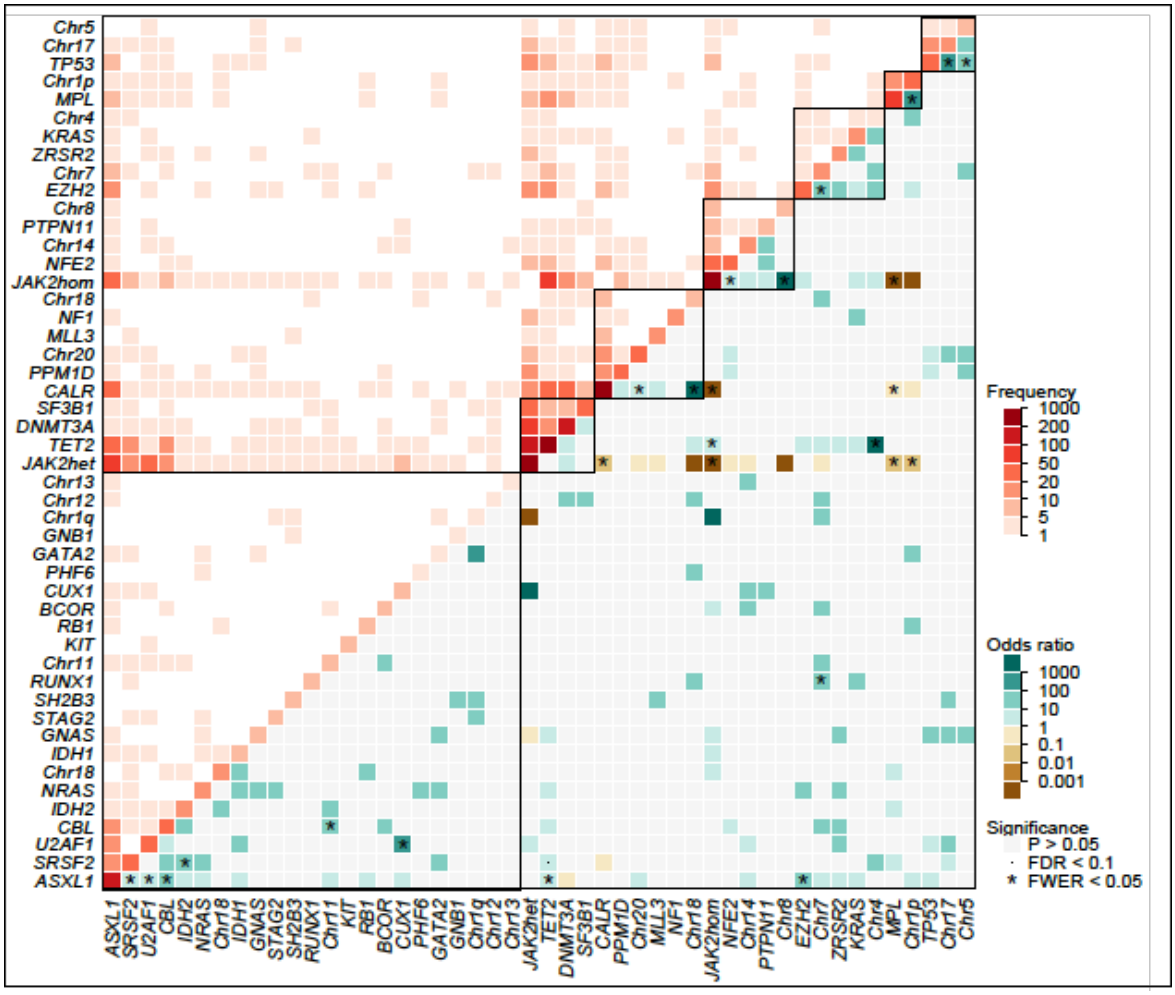


Figure 9: *Heatmap showing frequencies of mutation/chromosomal change co-occurrence (top-left triangle) and odds ratios (bottom right). Black squares represent the sets of genomic events most correlated with the 7 genomically-defined classes. FDR – false discovery rate, FWER – familywise error rate.*

The classes were therefore defined as follows:

- **MPN with TP53 disruption/aneuploidy** – defined as the presence of TP53 mutation, 17p del/UPD and/or 5q-. (2.5% of patients)

- **MPN with chromatin/spliceosome mutation** – in some iterations this class was separated into two classes defined predominantly by {SRSF2, U2AF1, NRAS, CBL, GNAS, IDH1/2, STAG2, CUX1, BCOR, RUNX1, PHF6} and {EZH2, KRAS, ZRSR2, 4qUPD, 7qUPD/del, ASXL1}, which appears to be related to the mutual exclusivity of EZH2 and SRSF2, seen both in this and other datasets. However, ASXL1, 7qUPD/del and ZRSR2 were not infrequently present in

patients categorised into the {SRSF2, U2AF1 etc.} subgroup (see also figure 9). For simplicity, therefore we have combined this into one group. (14.2%)

- **MPN with CALR mutation** – defined as the presence of CALR mutation (and therefore enriched for 19pUPD), and/or $20q^-$. (17.2%)

- **MPN with MPL mutation** - defined as the presence of MPL mutation, and enriched for 1pUPD. (2.6%)

- **MPN with homozygous JAK2 or NFE2 mutation** – defined as the presence of JAK2 mutation in conjunction with 9pUPD or trisomy 9 and/or NFE2 mutation. This group was additionally enriched for TET2 and PTPN11 mutations and trisomy 8. (13.5%)

- **MPN with heterozygous JAK2 mutation**. This was the commonest class – 38.4%

- **Myeloproliferation with no known driver mutation** (9.4%) or **Myeloproliferation with other driver mutation/unclassifiable** (2.2%).

3.4 Relative timing of individual mutations

We determined order of mutation acquisition for 271/671 patients harbouring multiple mutations using variant allele fractions corrected for copy-number change. Mutations in MPL, CALR, SF3B1 and DNMT3A were more frequently early events, while NRAS, KRAS, NFE2, PPM1D and TP53 were typically late events (Figure 10). Mutations in JAK2 were more variable, coming first in a given pair in 34% of cases.

We also assessed 330 patients at multiple disease time-points (median interval 7.7 years). Late acquired mutations were identified in 29 patients, suggesting in general the number of oncogenic mutations remains relatively stable following presentation. These 29 late-acquisition mutations involved PPM1D (n=10), TP53 (n=3), TET2 (n=7) and NFE2 (n=2), corroborating results from the single time-point analysis (Figure S3). Furthermore, patients with NFE2, TP53, KRAS, NRAS, ZRSR2, U2AF1, SRSF2 presented with these mutations over the age of 60 in 58, 59, 71, 72, 80, 82 and 89% of cases respectively, in contrast with only 33% of CALR-mutated patients.

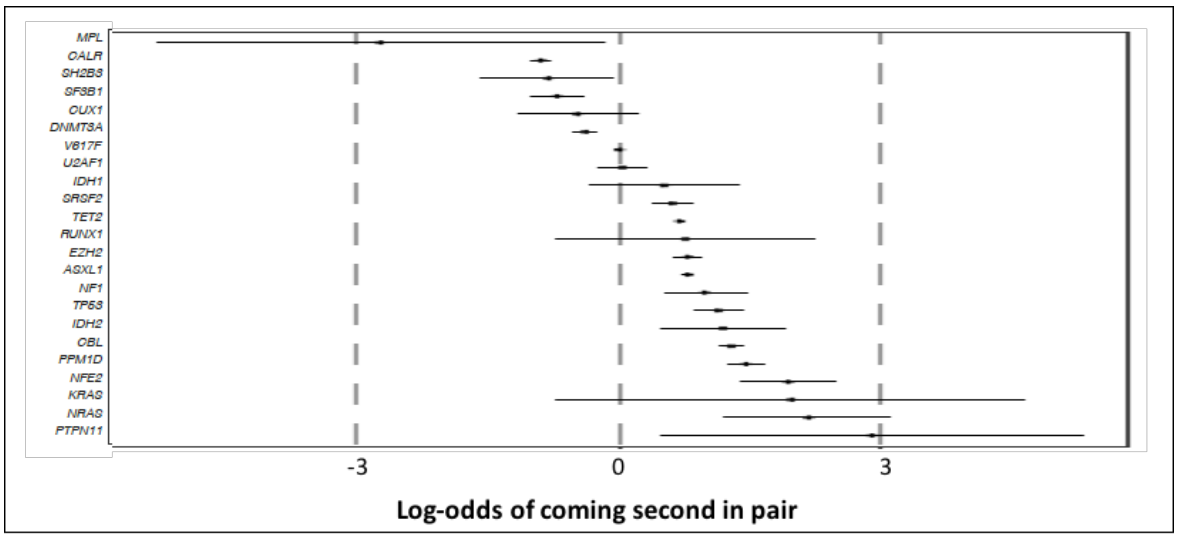


Figure 10: *Bradley–Terry model derived estimates of the overall timing of mutation acquisition. The horizontal axis shows the log odds of a gene occurring second in a gene pair. For example, as compared with JAK2(V617F), PPM1D mutations have a log odds of 1.45 and therefore are* $e^{1.45}$*=4.3 times more likely to occur second in the pair. Any pair of genes can be assessed in this manner by calculating the exponential of the difference in log odds for gene A and gene B. The error bars indicate quasivariances.*

3.5 Associations between germ-line variants and somatic variations

3.5.1 JAK2 46/1 haplotype

62 germ-line SNPs known to be associated with variation in haematological parameters and/or with risk of acquiring MPN or MPN-associated mutations (and were adequately covered across the entire cohort) were genotyped. Testing for associations between these SNPs and either somatic mutations or phenotype (discussed later) was carried out in a similar method to that described above. This data however carries the caveat that in some cases heterozygosity for an allele may erroneously be called as a homozygosity due to UPD, since for the vast majority of cases only tumour (full blood/granulocytes, rather than constitutional) samples were available. For this reason, only associations that held for presence/absence for an allele overall were investigated further, and, where possible the precise genotype was interrogated using adjacent heterozygous SNPs (which would similarly be expected to be affected by the presence of UPD).

Pile-up of the 46/1 tag SNP rs12340895 demonstrated the presence of the G allele in 70% of JAK2-mutated cases, 66% of MPL-mutated cases, 54% of CALR-mutated cases and 42% of triple-negative cases, in comparison to a prevalence of 42% in the general

population. Relative to the allelic distribution for 1000 genome patients (n=2504 in total, 503 from European cohorts), therefore, the G allele was enriched not only for JAK2-mutated cases (p-value<0.0001), but also for MPL- and CALR-mutated cases (p-values 0.005 and 0.02 respectively in comparison to European controls)[197,210]. Associations between JAK2 or MPL mutations and JAK2 46/1 haplotype have been previously reported but one study failed to show an association for CALR[229]. However, this study only included 22 sporadic CALR-mutated cases and 270 controls and therefore may have been insufficiently powered to detect an association.

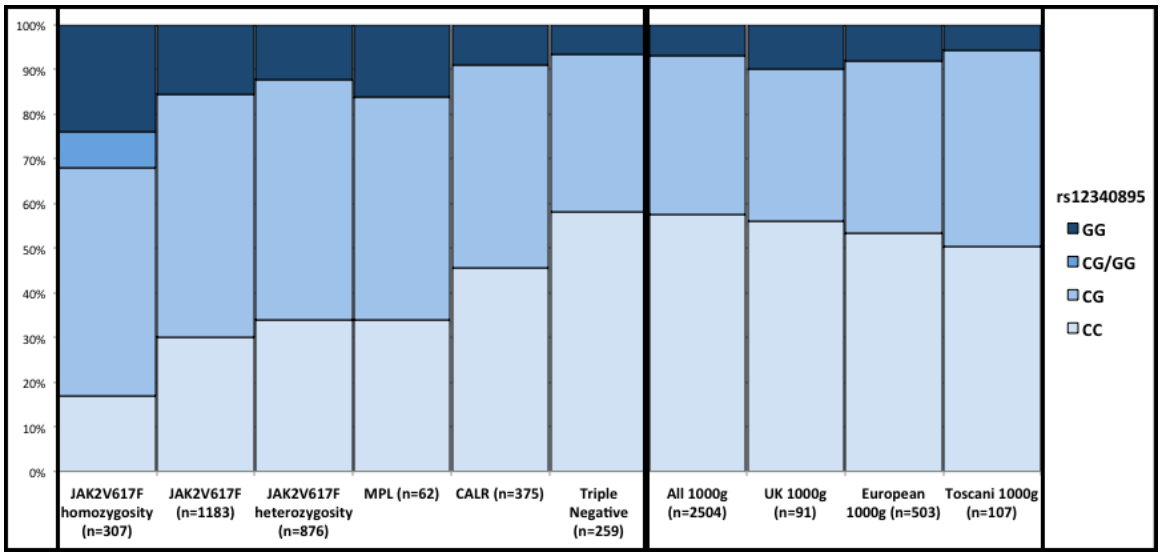


Figure 11: *Proportions of patients with GG/CG/CC genotypes of rs12340895 (corresponding to 46/1 haplotype) in this cohort (subdivided by phenotypic driver mutation) and control populations. 1000g=1000 genomes project*[197].

Further analysis restricted to JAK2-mutated cases additionally reveals a strong association between JAK2 haplotype and 9pUPD with prevalences of 15, 25-27 and 41-48% for cases with CC, CG and GG genotypes respectively (CG/GG genotypes cannot be definitely determined for 25 cases, which have complete loss of heterozygosity for 9p, since genotyping is carried out on blood samples alone). OR for 9pUPD and the presence of the G allele is 2.52 (95% CI: 1.8-3.6, p-value <0.0001) and 9pUPD occurs on the G allele chromosome in 92% of heterozygous (CG) cases. This result raises the possibility that 9p carrying the JAK2 46/1 haplotype has an increased likelihood of undergoing mitotic recombination resulting in loss of heterozygosity[230], that acquired homozygosity for a JAK2 mutation specifically in the context of the 46/1 haplotype provides an additional clonal advantage, or that homozygosity for the JAK2 haplotype itself provides an advantage in a JAK2-mutated context. However, no correlation was found between the

length of 9p affected by LOH or the size of the homozygous clone, and presence of the 46/1 haplotype. It should be noted that the association demonstrated here is not explained by independent bi-allelic acquisition of JAK2 in JAK2 46/1 homozygotes, although this occurrence is theoretically possible.

Collectively therefore, these results argue against hypermutability of the JAK2 46/1 haplotype (increasing the probability of JAK2V617F/exon 12 mutations) as the sole mechanism underlying its association with MPN susceptibility, suggesting that presence of the haplotype may increase the incidence of 9p LOH or affect its frequency, and provide a clonal advantage to, or indirectly influence the risk of development of, CALR- or MPL-mutated clones.

3.5.2 Other germ-line-somatic associations

A number of loci on chromosome 9p, beyond those defining the 46/1 haplotype, have been associated with haematological parameters, and therefore were genotyped in this study. Some of those assessed, including rs10974900, have high linkage disequilibrium with 46/1-defining SNPs and therefore share similar associations.

rs409801(T/C), an inter-genic variant telomeric to JAK2, has been correlated with platelet counts, with higher counts seen with the C allele[206]. The frequency of 9pUPD in JAK2-mutated patients was found to be positively correlated with the presence of the TT genotype (OR: 1.9, 95% CI: 1.4-2.5, p-value<0.0001), a correlation that remained significant after correction for JAK2 haplotype and diagnosis, and 9pUPD more commonly occurred on the T allele in heterozygous (CT) cases (59%).

rs741959(C/G), an inter-genic SNP located on chromosome 1p adjacent to TAL1[205], was found to correlate with the presence of 1pUPD. The GG genotype was present in 63% (12/19) of MPL mutated cases with 1pUPD, compared to 33% of MPL unmutated-cases or heterozygous MPL-mutated cases, as well as in the 2 MPL-unmutated cases displaying 1pUPD (in one case in the presence of the MPL R592Q variant). However, this association may still be spurious, since the observed fractions of both CC and GG genotypes are increased in cases of 1pUPD.

As with the case of the 46/1 haplotype, these cases hint at a clonal advantage arising from homozygosity for particular germ-line variants, or towards an influence of these variants on the risk of phenotypic driver mutation or mitotic recombination leading to LOH

occurring.

3.6 Discussion and Future work

This large-scale sequencing study demonstrates recurrent mutations in a set of 33 genes, as well as chromosomal gains, losses and copy-number neutral loss of heterozygosity across 14 chromosomes. The frequencies of these mutations and chromosomal changes, when ranked, fall rapidly such that only 4 genes were recurrently mutated in more than 5% and over half are seen in <1% of cases. This long tail of mutations reflects previous sequencing studies seen in MPN (Figure 12A) other myeloid malignancies, and well as other cancers, and in fact follows a distribution similar to that described by Zipf's law[231] (Figure 12B). Although the data only approximate to a power law, it is intriguing to note that the curve for MPNs falls more steeply than other malignancies, a relationship that is held across this dataset and data from COSMIC[203], with a similarly steep slope seen with CML.

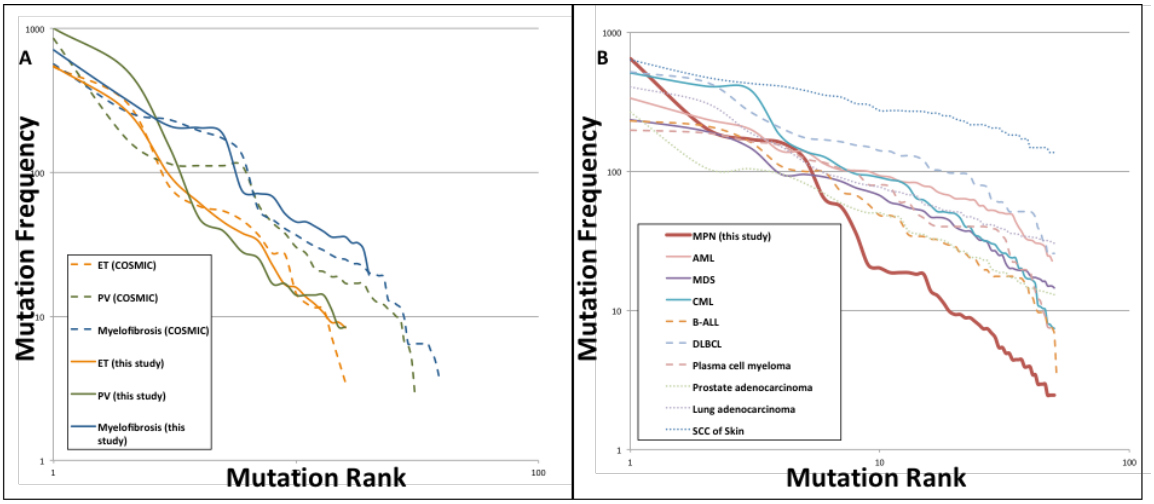


Figure 12: *Frequency of mutations plotted against mutation rank (by frequency) plotted on log-log scale, for different MPN subtypes as well as for other haematological and solid organ malignancies.*

This likely reflects the fact that MPNs can be driven by a small pool of phenotypic driver mutations (predominantly JAK2 and CALR) that are found in the majority of cases, and that for a given individual only a small number of mutations are detected (a median of 1-2). This argues against the requirement for additional mutations to counteract the lack of a stem cell advantage of a JAK2- or CALR-mutated clone in order to sustain disease; just over 50% of patients in this cohort were found to have an isolated JAK2, CALR or MPL mutation. However, the mutational burden of JAK2-homozygous cases is greater than that of JAK2-heterozygotes, with enrichment for TET2 and NFE2 mutations, suggesting this

hypothesis may still hold in the case of homozygous JAK2. Despite this, it does not appear that the development of an MPN, as opposed, for example, to JAK2-mutant CHIP, is determined simply by the presence/absence of additional (exonic) mutations. Further population-wide studies are required to determine genomic or environmental factors that differentiate JAK2-mutant CHIP from isolated JAK2-mutant MPN. While the presence of the JAK2 46/1 haplotype was shown to correlate strongly with the presence of JAK2 mutations, and also homozygosity for these lesions.

This sequencing study has not revealed any clear novel phenotypic driver mutations that serve a similar role to JAK2, CALR or MPL in that they are frequently shown to be the sole molecular abnormality associated with clonal expansion and a disease phenotype. While SH2B3 is often reported as a phenotypic driver, sufficient in its own right to induce disease[72], it was only seen in the absence of JAK2, CALR or MPL in one patient with ET, where it was found together with an MLL3 mutation. MLL3 and PPM1D are often found in conjunction with JAK2, CALR or MPL mutations, but were found in their absence in 10 patients (and in isolation in 8), suggesting they may be sufficient to drive clonal expansion and result in an MPN phenotype. Validation of these findings is required in an independent cohort of MPN patients, and if confirmed to be recurrent, functional studies are needed to investigate the effects of MLL3 mutations on cell biology, in particular on stem cell expansion and differentiation. Further analysis of other sequenced genes (those only sequenced in either TGS1 or TGS2) and of the results of the whole genome sequencing data may also reveal further candidate driver mutations which, together with the candidates mentioned in Sections 3.2.7 and 3.2.8 (including ATRX, EP300, CREBBP, IKZF2 and NRD2) could be sequenced in future MPN patient cohorts. Unfortunately, *Ikaros* (*IKZF1*) was not well covered by either TGS1 or TGS2 panels, and should be included on further sequencing bait sets to determine whether mutations or microdeletions are present.

Mutations in the panel of 33 genes analysed display distinct patterns of acquisition. Firstly, certain mutations tend to occur more often as initiating events, being present in the dominant clone in most cases. These include MPL and CALR, confirming previous results[52,232]. Others, including NRAS, NFE2, TP53 and PPM1D occur most often as secondary events. Furthermore, distinct patterns of co-mutation, clustering or mutual exclusivity were identified. JAK2, CALR and MPL were shown to be mutually exclusive (in all but a very small number of cases), reflecting both their sufficiency for generating a disease phenotype and functional redundancy.

Mutual exclusivity was also observed for EZH2, SRSF2 and U2AF1 (even though each of these mutations independently was enriched in ASXL1 mutated patients), a finding also seen in patients with AML and MDS[106,189,214]. This hints at functional redundancy again, and this is supported their similarities in co-mutated genes and by studies reporting reduced EZH2 expression in SRSF2-mutated cases, and demonstrated reduced H3K27 methylation not only as a logical consequence of EZH2 mutation, but also in SRSF2- and U2AF1-mutated cases[100,106]. The phenotypic correlates of these mutations are discussed in a subsequent section, but work further exploring the overlap in epigenetic and transcriptional changes as a result of these lesions and their effects on cell biology, as well as the potential for functional synergy with ASXL1 mutations would be informative, particularly given their co-occurrence across MPNs, MDS and AML.

4. Impact of genomic variation on MPN phenotype

4.1 Introduction

Alterations in cytokine signalling pathways play a crucial role in the development and phenotype of myeloproliferative neoplasms. MPL and CALR mutations are understood to predominantly disrupt the thrombopoietin pathway and for this reason are found almost exclusively in patients with ET or MF, rather than PV[47,52,51]. In contrast, patients with PV exclusively carry mutations in JAK2, disrupting erythropoietin, as well as thrombopoietin and G-CSF, signalling pathways, and patients with PV are more likely to have mutations in exon 12, harbour a dominant JAK2 homozygous subclone and/or to have the JAK2 mutation preceding a TET2/DNMT3A mutation[26,40,122,123]. A handful of mutations (including SRSF2, ASXL1 and EZH2) have been found more commonly in MF, and are associated with a poorer prognosis[69,233,234]. Beyond this however, clinical heterogeneity remains poorly explained, both in terms of presenting diagnosis (ET, PV or PMF) and, particularly in chronic phase disease, the risk of subsequent transformation to MF or AML. The comprehensive genomic analysis performed on this large cohort enables to further explore the associations between somatic and germ-line factors and diagnostic classification and transformation risk.

More broadly speaking, the diagnostic categories of ET, PV and MF, are defined according to clinical, histological and genomic factors, but it is recognised that several of these features overlap across the three as well as with MDS/MPN overlap syndromes, MDS and AML. Therefore, an unbiased classification approach may offer novel methods for disease classification, potentially with reduced reliance on subjective criteria and greater reproducibility.

4.2 Phenotypic associations of genomic variables

4.2.1 Associations between mutations and subtype – MF vs. CP disease

Using frequent pattern mining, pairwise odds ratios and significance testing, as described in sections 3.3.1-3.3.2 a number of somatic genomic changes were found to correlate with myelofibrosis rather than chronic phase disease (ET or PV) at the time of sampling. As expected, mutations of chromatin regulators and spliceosomal components, as well a number affecting signalling pathways (CBL, GNAS, NRAS, and 9p and 1p UPD leading to LOH for JAK2 and MPL mutations) were strongly enriched in patients with MF.

However, as noted above, a number of these mutations are commonly co-mutated, but are also more common in older patients. Logistic regression was therefore carried out to determine which variables, including germ-line variants, independently predicted for MF phenotype once age and gender are also taken into account. The results of this analysis are show in Figure 13A and confirm that the majority of these mutations retain independent associations with MF, together with increasing age, male sex and a small number of tested germ-line variants.

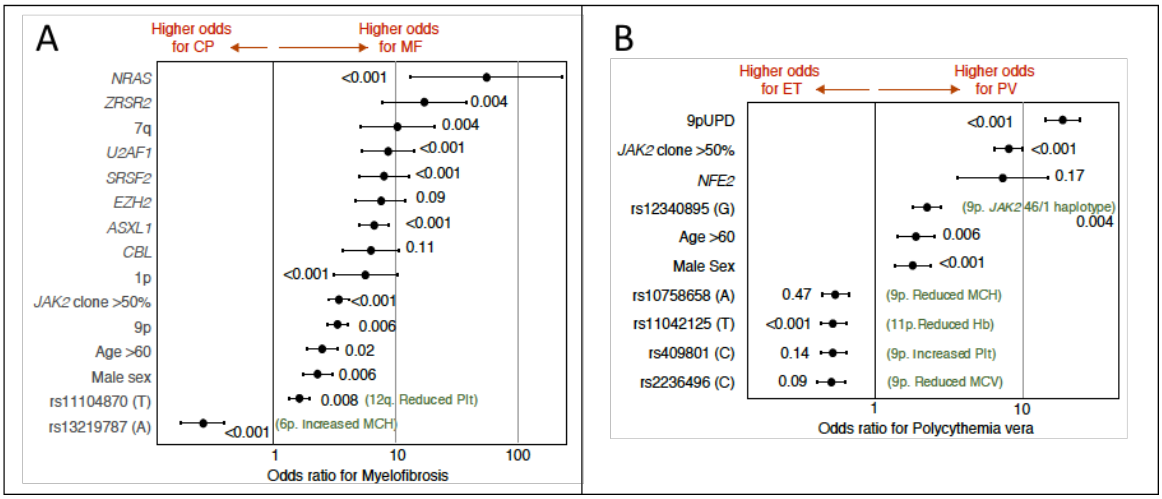


Figure 13: *Odds ratios associated with each variable for (A) myelofibrosis compared to chronic phase disease across whole cohort, and (B) PV compared to ET for JAK2 mutated chronic phase patients. P-values supplied derived from logistic regression modelling using all candidate predictors (including age as continuous variable).*

rs11104870(C/T) is an inter-genic SNP on chromosome located upstream of *KITLG*, which encodes KIT-ligand or stem cell factor (SCF), and shown to affect its expression, has been correlated with red blood and platelet counts in genome-wide association studies[205]. Presence of the T allele was associated with an MF phenotype (Figure 14A, OR 2.0, 95% CI 1.1-4.0 for CT genotype, OR 2.9 95% CI 1.6-5.6 for TT genotype, p-value=0.008 in multivariate analysis). The role of genomic variables in determining the risk of transition between disease states is addressed more comprehensively in Section 4.4, but intriguingly, rs11104870 genotype alone correlated with rates of myelofibrotic transformation from ET or PV, an association that remained significant after adjusting for age and gender (Figure 14B, HR for number of T alleles 2.13, 95% confidence interval: 1.25- 3.63, p-value=0.005). As noted above, it is possible that this association is confounded by LOH at this locus, however only 4 cases of 12qLOH were identified in this cohort (1 genotyped as CT, 3 as TT).

rs13219787(A/G) is located on chromosome 6p within the histone 1 cluster and is shown to correlate with haemoglobin and mean cell volume[205]. Presence of the A allele was less common in patients with MF, with the GG genotype having an odds ratio of 3.9 for MF (95% CI: 2.1-8.0, p-value<0.0001 in multivariate analysis, Figure 14C), and was associated with an average increase in haemoglobin of 2.8g/l in linear regression analysis (p-value=0.04).

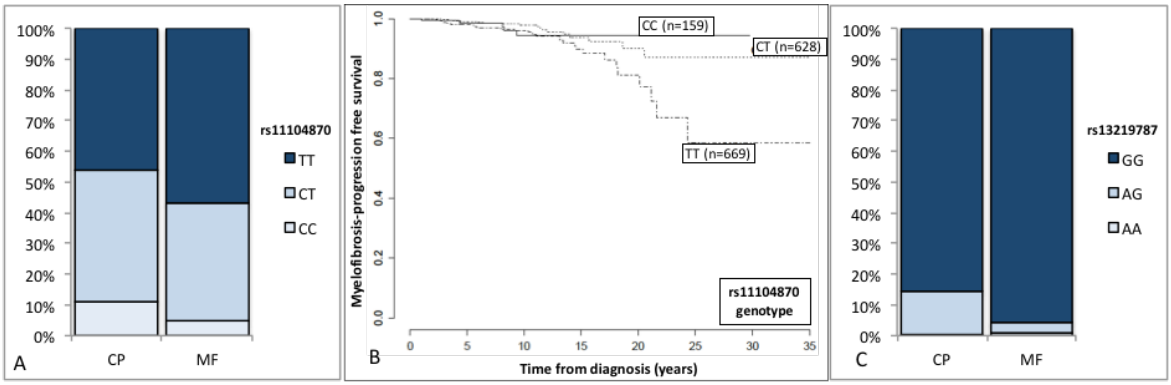


Figure 14: *Correlation of germ-line SNPs and MF phenotype. (A) rs11104870 genotype and MF phenotype at time of sample, (B) rs11104870 and MF transformation from chronic phase, (C) rs13219787 genotype and MF phenotype at time of sample.*

4.2.2 Associations between mutations and subtype – ET vs. PV

Arguably more intriguing are the mechanisms underlying the manifestation of a PV rather than an ET phenotype in JAK2V617F-mutated patients. A similar analysis to that described in 4.2.1 was performed, with analysis restricted to JAK2V617F-mutated patients with either PV of ET (Figures 13B and 15). Unsurprisingly, as reported in previous studies, both JAK2V617F clone size (arbitrarily dichotomised at >50%, giving OR 8.1, 95% CI: 5.9-11.3) and the presence of 9pLOH (OR 18.6, 95% CI: 12.6-28.1) correlated independently with a PV phenotype.

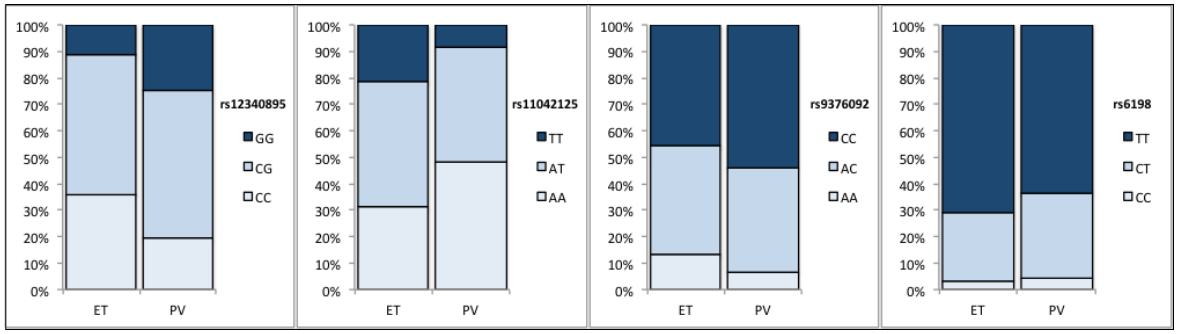


Figure 15: *Allele frequencies for rs12340895, rs11042125, rs9376092 and rs6198 across patients with PV or JAK2-mutated ET.*

The 46/1 haplotype also correlated with the PV phenotype (p-value<0.0001); an association independent of 9pLOH in multivariate analysis and as such was maintained when analysis was restricted to patients without detectable 9pLOH (p-value=0.005), but not observed in those with 9pLOH (p-value=0.48). This further supports a role for the 46/1 haplotype in MPN pathogenesis beyond its association with the presence of JAK2V617F alone. It is also worth noting that although numbers of JAK2-homozygous patients with ET are limited (n=44), and the resolution of the assessment of 9pLOH is also low, the length of chromosome involved (as determined by allele frequencies along 9p) appears to be shorter in ET than in PV or MF, but appeared to independent of JAK2 46/1 status (Figure 16).

All six *STAG2*- and JAK2V617F-mutated patients had a PV phenotype, but this association did not reach statistical significance after correction for multiple hypothesis testing. *NFE2* mutations and rs409801(T/C) also were enriched in patients with PV, but these associations did not reach significance in multivariate analysis, likely reflecting the co-occurrence of these variables with 9pLOH. rs11042125(T/A) however strongly correlated with phenotype[205]. This SNP is located on Chr11p and has been associated with a number of genes (AKIP1, C11orf16, NRIP3, ST5). The T allele has been correlated with lower haemoglobin concentrations in GWA studies, and in keeping with this is less frequently seen in patients with PV, with the TT genotype seen in in 21% of ET but only 8% of PV cases (p-value <0.001 for the T allele). The T allele was associated with, on average, an increase in platelet count of $51x10^9$/l and decrease in haemoglobin of 3g/l in linear regression analysis within this cohort (p-value=0.001).

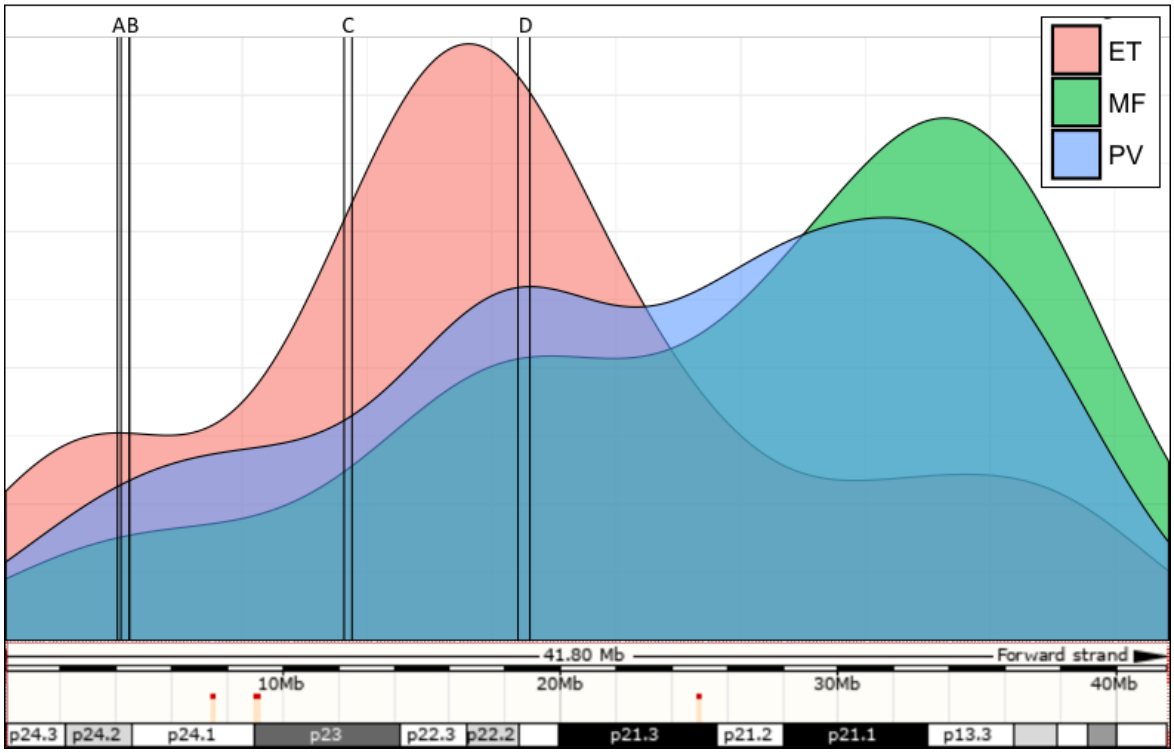


Figure 16: *Density plot showing lengths (in base pairs) involved by 9pLOH, estimated using heterozygous SNP allele fractions in 44 ET, 174 PV and 92 MF cases. A denotes JAK2, B: CD274 (PD-L1), C: NFIB, D: interferon cluster.*

Previous studies have highlighted two other SNPs as potentially playing a role in determining PV vs. ET phenotype (Figure 15). Variants in the HBS1L-MYB inter-genic region have been correlated with MYB expression and with a number of haematological variables. One of these, rs9376092, was found to significantly correlate with an ET phenotype in JAK2-mutated patients[140]. In this analysis, the number of A alleles correlated with the probability of having an ET phenotype (p-value=0.0019), but did not remain a significant predictor in logistic regression including the other variables mentioned above.

Finally, SNPs in the glucocorticoid receptor (GR) have previously linked to steroid responsiveness as well as alterations in the regulation of erythropoiesis. They therefore present attractive targets as a germ-line predictors for PV, and have been shown in some studies to play a role in this[148,149]. In this cohort, rs6198, associated with expression of a dominant negative isoform of GR-β, did associate with phenotype (p-value=0.019), but this did not reach statistical significance after correction for multiple hypothesis testing, or in logistic regression analysis. However, these findings do not preclude a role for these variants in certain contexts.

4.2.3 Associations with haematological parameters

Linear regression analysis was performed to determine which variables were the greatest determinants of haematological parameters, taking into account age and gender. Increasing age was associated with a decline in haemoglobin concentrations and rise in white cell counts, while male sex was associated with higher haemoglobin and lower platelet counts.

Mutations associated with an MF phenotype (in ASXL1, NRAS and components of the spliceosome), as expected, were associated with lower haemoglobin concentrations, but also in some cases also with lower platelet counts (NRAS and U2AF1) or raised white counts (SRSF2). TET2 mutations were associated with higher white cell counts, but particularly in cases of TET2 homozygosity. These findings illustrate that even for a given MPN subtype , the individual combination of demographic and genomic variables can fine-tune and individual's phenotype.

Table 7: *Estimated change in haemoglobin (g/l), white cell count ($x10^9$/l) and platelet count ($x10^9$/l) associated with each variable, derived from linear regression modelling.*

	Haemoglobin	White cell count	Platelets
Male sex	8.28		-68.69
JAK2 V617F	10.72		
9p	11.58	3.59	-226.63
Age (per decade)	-0.98	0.35	
SRSF2	-18.67	3.55	
4q	-30.96	23.48	
NRAS	-26.98		-420.32
U2AF1	-30.36		-176.64
ASXL1	-9.94		
SF3B1	-10.78		
ZRSR2	-23.96		
Trisomy 8	-38.72		
5q		-7.09	
CALR		-1.65	192.51
TET2		1.39	
17p		5.56	
Trisomy 9		5.87	
GATA2		14.96	
CBL			-232.09
7q			-325.61
JAK2 exon 12			-411.81

4.2.4 Phenotypic correlates of genomically defined sub-groups

As would be expected from the above analysis, certain genomically-defined groups (as defined in **3.3.3**) show enrichment for different phenotypes (Figure 17). In particular, the homozygous JAK2 or NFE2 mutation group is enriched for patients with PV, while the chromatin/spliceosome mutation group is enriched for patients with MF and the heterozygous JAK2 mutation group is more commonly associated with chronic phase disease rather than myelofibrosis (p-values <0.0001 in all cases).

However, beyond this, genomically-defined groups have a number of features that transcend MPN subtype, suggesting that this genomic classification is able to pick out groups of patients with phenotypic similarities.

The chromatin/spliceosome mutation group is associated with poorer overall survival, whether patients are classified as having myelofibrosis (p-value=0.004, with JAK2-heterozygous group as comparator for all analyses) or chronic phase disease (p-value<0.0001), and additionally picks out a subset of patients with chronic phase disease with higher risk of myelofibrotic or AML transformation (HR 5.4, p-value<0.0001 for each outcome). These findings are confirmed in an independent cohort of 151 CP patients (with sufficient genomic data to allow classification) and 190 MF patients: p-values=0.02 for transformation from CP and <0.0001 for event-free survival in MF)

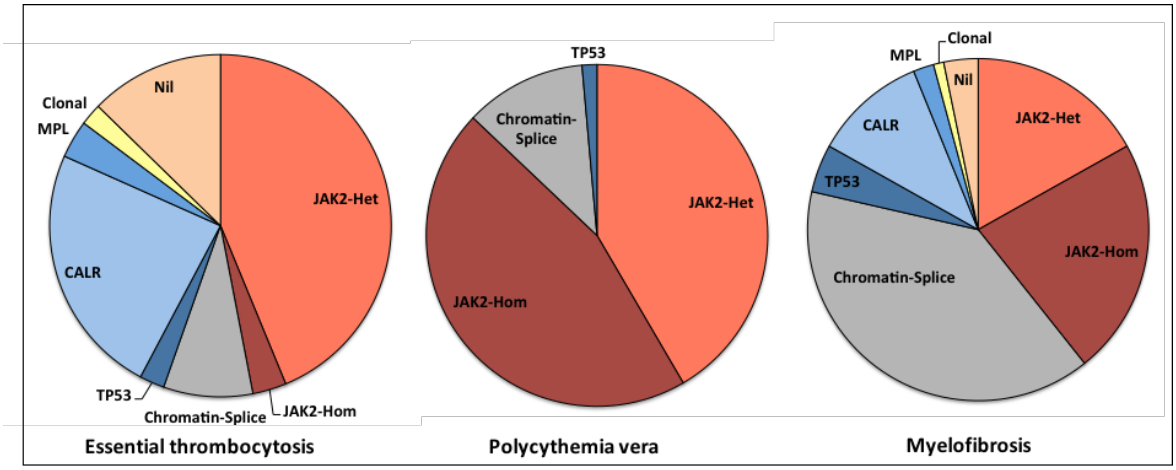


Figure 17: *Proportions of each genomically defined subgroup in patients with Essential thrombocytosis, Polycythemia vera and myelofibrosis*

The TP53 disruption/aneuploidy group is significantly associated with progression to AML regardless of MPN subtype (HR 10.4, p-value <0.0001 for chronic phase disease, HR 15.8, p-value=0.0006 in myelofibrosis), and highlights a subset of patients with myelofibrosis with poorer overall survival (HR 2.6, p-value=0.002). This was confirmed

in MF for the external validation cohort (p-value=0.002), but there were insufficient CP patients tested for and found to have a TP53 mutation in the external cohort to evaluate this association.

In contrast, the group with no known driver mutation, appeared to carry a low risk of disease transformation or death, with only 2 AML and 1 MF transformation event observed during the follow-up of 183 patients in this group (p-values 0.0004 and 0.0023 in univariate analysis). However, these patients were significantly younger (median age 41.5 years, compared to 58 years in the remaining groups, p-value<0.0001), and tended to be female (70.3% compared to 51.8% in the remaining groups, p-value<0.0001), and this prognostic association was not independent of age and sex. Despite this, this group appears to highlight a subset of patients with a more benign phenotype. Although in all cases there was convincing histological evidence for an MPN, a non-clonal cause for the MPN-like phenotype in these patients cannot be entirely ruled out in the absence of further clonality studies.

It is also worth noting that Cox proportional hazards models using the Bayesian classes defined above alone, are as effective (in terms of correctly ranking patients by risk, i.e. their concordance, as discussed in more detail in **Sections 2.8.6 and 5.9**) as models that categorise patients according to MPN subtype (ET/PV/MF), with concordances of 64% respectively, and are superior in their ability to predict AML transformation (79 vs 75%).

4.3 Clinical correlates of mutation order and clonal composition

4.3.1 Phenotypic driver acquisition order and phenotype

The above analysis addresses the correlates of specific mutations (as well as chromosomal changes and germ-line SNPs) and combinations thereof. However beyond the presence/absence of mutations, it is recognised that the order in which the mutations occur in can play a role in influencing disease phenotype, adding an extra layer of complexity into the association between genotype and phenotype. The presence of 2 mutations may correspond to 3 different orderings/clonal structures, of only 3 mutations to 16 possible orderings and of 4 to 125 possible orderings[235]. As the number of mutations increases the number of possible clonal structures increases super-exponentially (figure 18), which may in part explain how clinical heterogeneity outpaces genomic heterogeneity. Therefore, although MPNs are generally genomically simple, since only ~40% have 2 or more mutations, the order of these mutations may still play a significant role in determining

phenotype.

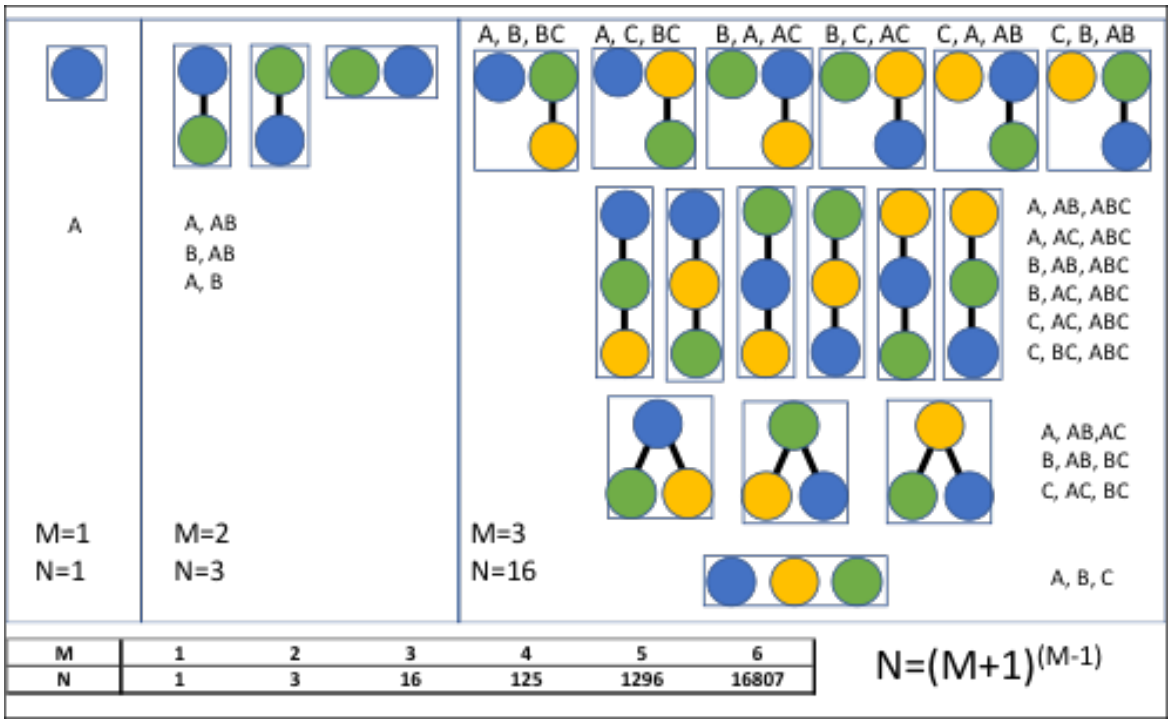


Figure 18: *Possible clonal structures in cases where 1, 2 and 3 mutations are detected for mutations denoted A, B and C.*

As noted above (**sections 3.4.1-3.4.2**), CALR and MPL tended to be the earliest mutational events when present in combination. In patients with multiple mutations, the timing of JAK2V617F acquisition was more variable. As mutation order of JAK2V617F relative to DNMT3A and TET2 mutations has been shown to correlate with MPN phenotype[122,123], we assessed the effect of timing of acquisition of JAK2V617F globally. Of 480 JAK2V617F-mutated patients (as the sole phenotypic driver) with additional driver mutations, order could be evaluated in 140 (29.2%). JAK2V617F was preceded by at least one other mutation, most commonly in DNMT3A, TET2, ASXL1 or SF3B1, in the majority of patients with ET (74.4% - figure 19A). In contrast, JAK2V617F mutations were the earliest detectable event in most patients with PV or MF (70.5 and 62.3% respectively).

9pUPD was detectable in 62.7% of JAK2-first cases but only in 18.5% of JAK2 second cases (p-value<0.0001). However, the association between JAK2-first cases and JAK2-homozygosity does not appear to be the sole mechanism underlying the phenotypic differences: when analysis is restricted to cases without 9pUPD there is still a trend towards JAK2-second status in ET (Figure 19B, p-value=0.006).

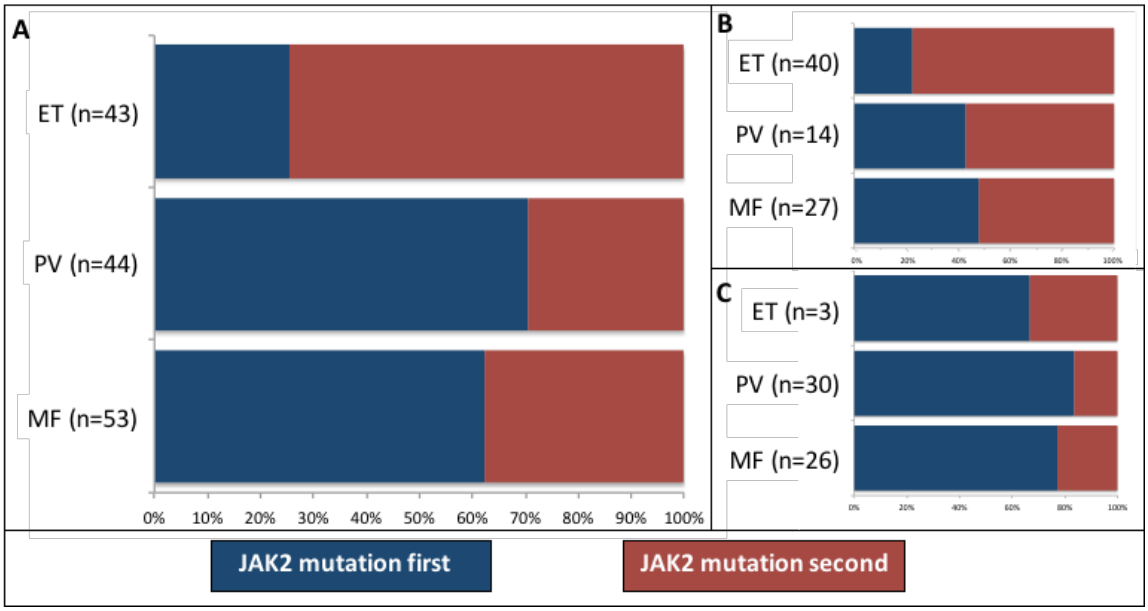


Figure 19: Associations of phenotype (ET, PV, MF) with order of JAK2 mutation acquisition relative to partner mutations. (A) All JAK2-mutated patients with multiple mutations for whom order could be determined, (B) Subset of A without detectable 9pUPD. (C) Subset of A with detectable9pUPD.

This analysis includes relatively few patients but suggests some further potentially interesting associations between not only phenotype and the relative timing of JAK2 mutations, but also with the specific additional mutated gene.

Firstly, correlation of JAK2-first and PV and JAK2-second with ET appears to hold regardless of partner mutation, and this confirms previous reports regarding JAK2 vs. TET2 or DNMT3A, but also expands it to other partner mutations (Figure 20A and B). However, this does not appear to be the case for patients with MF, for whom there appears to be enrichment for JAK2 mutation-first, ASXL/EZH2 mutation-second status (Figure 20C), while the timing of JAK2 relative to TET2/DNMT3A mutations follows a similar pattern to patients with ET.

Secondly, the clonal hierarchy of JAK2 mutation acquisition following a DNMT3A or SF3B1 mutation appears to be have high specificity for ET: JAK2 was preceded by DNMT3A or SF3B1 in 23.3% ET cases in this subgroup analysis (making up 31.2% of all JAK2-second ET cases), compared to 3.7% of MF and 0% of PV. Together, these analysis suggest that whether JAK2 occurs as a primary or secondary event is not the only feature determining phenotype, but that the phenotypic associations also vary depending on partner mutation.

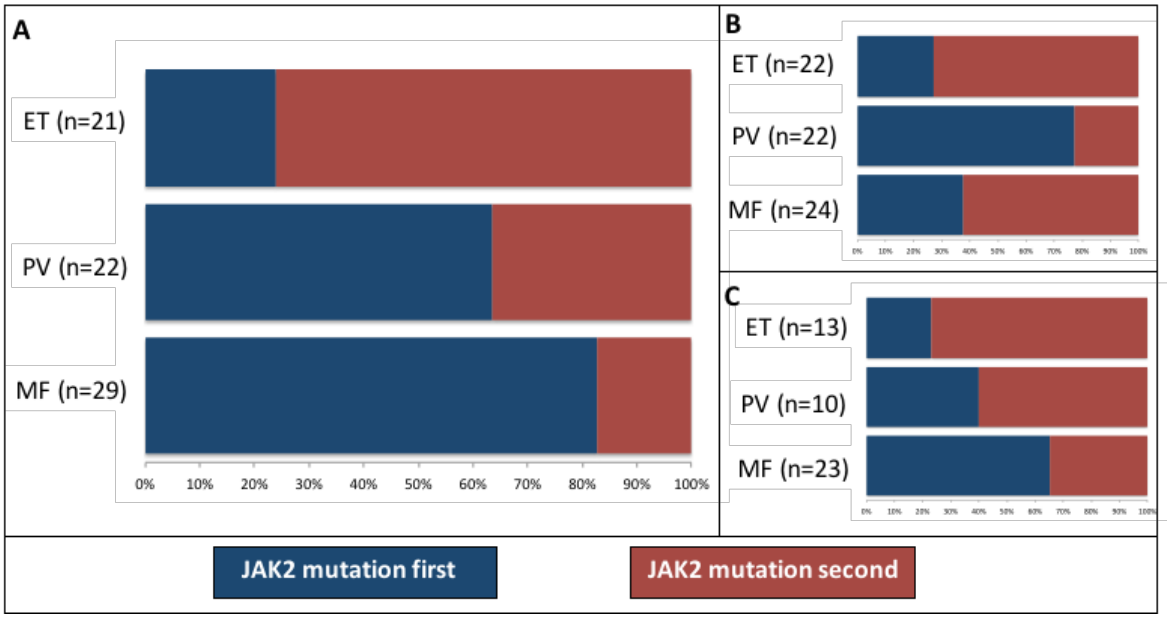


Figure 20: *A. JAK2 patients with multiple mutations, compared to non-TET2 or DNMT3A mutations. B. TET2 and DNMT3A mutated cases only. C. EZH2 and ASXL1 mutated cases only.*

4.3.2 Clonal composition and phenotype

One limitation of the above analysis is that it is restricted to patients with sufficiently high clonal fractions for each of the two mutations being compared and requires a significant difference in these fractions. It therefore does not include approximately 70% of the patients with multiple mutations who have either too low clonal burdens or similar fractions for each of the two mutations, or both.

We therefore extended our analysis of JAK2-mutated patients more broadly by classifying patients into those where (a) the JAK2V617F is in the largest/dominant clone and the clone carrying the partner mutation is significantly smaller (red in figure 21), (b) JAK2V617F and the partner mutation are both in the dominant clone but no significant difference in clone sizes can be determined and therefore order of acquisition is undetermined (dark red) (c) JAK2V617F and the partner mutation are found in the same clone, but JAK2V617F is subclonal (dark blue), (d) the clone sizes of JAK2V617F and the partner mutation are such that biclonality cannot be excluded (light blue).

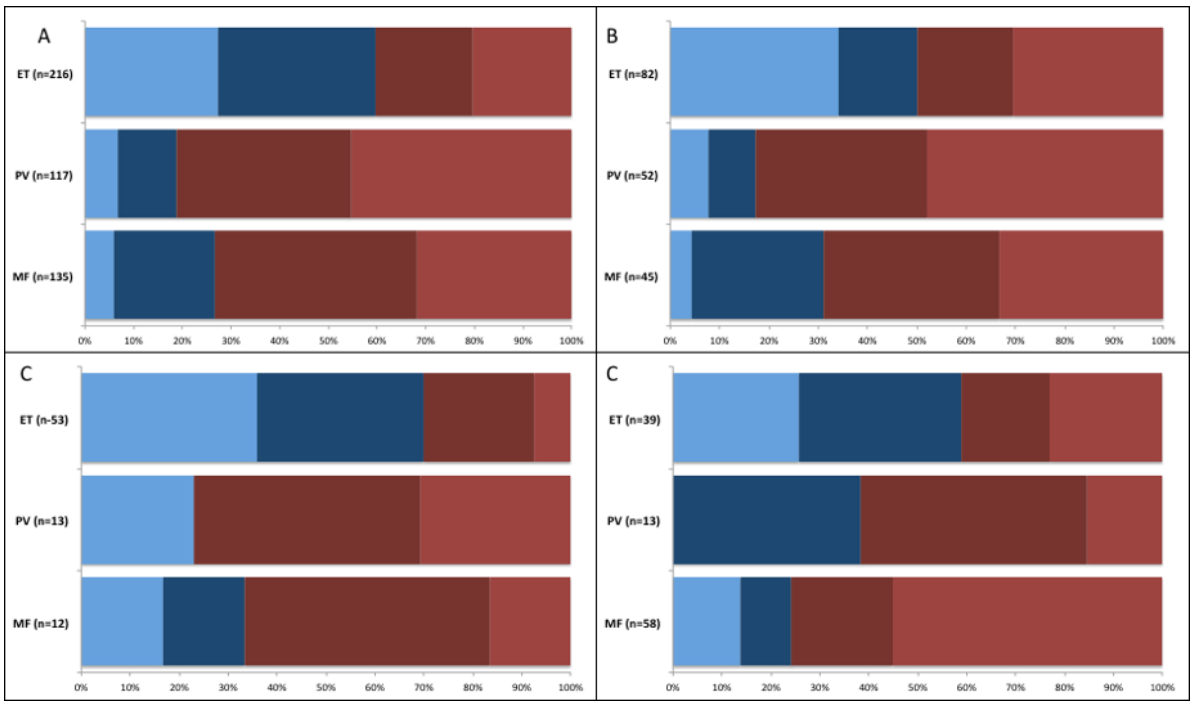


Figure 21: *Clonal composition for (A) All JAK2 mutated patients (B) JAK2 relative to TET2, (C) JAK2 relative to DNMT3A, (D) JAK2 relative to ASXL1/EZH2. Colouring as documented in text.*

The results of this analysis largely confirmed those of the previous one. JAK2V617F was most frequently sub-clonal in ET (59.7%), but this included a much higher proportion of cases (27.3%) where both clones were small, order could not be determined and bi-clonality could not be excluded. JAK2V617F occurred in the dominant clone (either with or without additional mutations) more commonly in PV and MF (81.2 and 73.3% respectively, $p<0.0001$)

The ordering of JAK2 mutation acquisition relative to specific partner mutations was again confirmed as correlating with phenotype in this wider analysis. Specifically, subclonal JAK2 in the context of TET2 or DNMT3A mutations was much more common in ET than PV (p-values 0.0002 and 0.003 respectively), and ASXL1 and EZH2 were more frequently shown to be sub-clonal events in JAK2-mutated MF than ET or PV (55 vs 21%, p-value 0.0005).

4.4 Discussion and Future work

Despite the relative genomic simplicity of MPNs, somatic mutations are here shown to significantly contribute to disease phenotype and natural history, and themselves demarcate distinct entities within the MPNs. These genomic entities show clear

phenotypic differences, and while not fully accounting for phenotypic heterogeneity across the MPNs, genomic classification offers a number of advantages over the traditional division into ET, PV and MF. In particular, genomic criteria are objective and reproducible, and to some extent reflect underlying disease biology, a feature that not only has the potential to further our understanding of disease mechanisms but may also facilitate the development or delivery of targeted therapeutic agents.

Here we show that the phenotypic differences between chronic phase disease (ET/PV) and MF are strongly associated with:

(a) Mutations in a cluster of genes that correspond to the genomic group referred to above as the "chromatin/spliceosome" group and which include ASXL1 and EZH2, SRSF2, U2AF1 and ZRSR2. As discussed previously, there is some evidence that several of these mutations affect PRC2 function and H3K27 methylation and affect stem cell function[100]. Further work is required to elucidate the effects of these mutations on down-steam transcription and cell biology.

(b) Mutations or chromosomal changes associated with increased cytokine receptor pathway signalling, namely 1pUPD (in the context of MPL mutations), 9pUPD (in the context of JAK2 mutations), larger JAK2 clones, CBL and NRAS mutations.

(c) Germ-line differences, including sex and a number of SNPs, including one associated with KIT ligand.

The mechanisms underlying these associations remain unclear. One possibility is that the characteristic processes underlying the myelofibrotic phenotype, namely bone marrow fibrosis with resultant bone marrow failure, cytopenias and extra-medullary haematopoiesis, are natural consequences of myeloproliferation in the marrow, due for example, to increased secretion of pro-inflammatory, pro-fibrotic and pro-angiogenic signals, increase reactive oxygen species and bone marrow remodelling. This is in keeping with the correlation of myelofibrosis with increasing age. In this scenario, this process is occurring at different rates in different patients (with ET or PV, or in the absence of a clinically apparent MPN), and the presence of the factors described above (in a-c) either serves to accelerate the process or leads to an earlier presentation of MF due to more rapid development of cytopenias. Specifically, the set of mutations described in (a) are shown

here to be associated with anaemia, and in some cases with thrombocytopenia, and this may due to a direct haematopoietic defect, a hypothesis supported by their prevalence in MDS.

A second (although not mutually exclusive) possibility is that these mutations (or germ-line factors) are directly responsible for alterations in tumour biology that result in novel behaviour, which may include alterations in secretions of soluble mediators, that in turn drive fibrosis or angiogenesis. This in turn is what leads to the cytopenias seen with these mutations rather than a direct effect of the mutations on haematopoiesis. In order to test this hypothesis, more work is needed to determine which soluble factors correlate with the development/presence of fibrosis, and whether their secretion is correlated with particular patterns of mutation. This is further explored in chapter 6.

We also explore the phenomenon wherein patients with the same point mutation in JAK2 can have either an ET or PV phenotype. As is previously reported, PV is strongly correlated with the presence of a JAK2-homozygous clone. NFE2 mutations appear to be enriched in patients with homozygous JAK2, and therefore in PV, suggesting a possible synergy between the two. This is in keeping with the association of NFE2 overexpression with erythropoietin independence and increased erythropoiesis[114]. In addition to correlating with the presence of JAK2V617F, the 46/1 haplotype correlated with the presence of 9pUPD and independently with PV. Whether the presence of 46/1 haplotype itself increases the risk of developing 9pUPD or 9pUPD causing homozygosity for the haplotype has an additional advantageous role for the clone (or both) remains to be seen. Also intriguing are the findings that the length of 9pUPD, where present, appears to correlate with ET vs PV phenotype (discussed later), and the genotypes of germ-line SNPs (outside of 9p and correlated with blood counts in wider studies) also appear to influence the overall MPN phenotype. These findings also require confirmation in further studies, both in terms of larger scale, more comprehensive identification of germ-line variation (with particular focus on 9p), but also the functional consequences of these variants.

This study also confirms, on a larger scale, the finding that patients with ET are more likely to have acquired JAK2 as a secondary event (for example, as a subclonal event within a TET2- or DNMT3A- mutated clone). A number of potential mechanisms may underlie these phenotypic associations, including cell-intrinsic differences in double-mutant clones depending on mutation order, and alterations in the microenvironment induced by the first clone. Furthermore, in JAK2-first patients, the JAK2 mutation will be

present in all mutant cells, and (except in some cases with >2 mutations) will not be in competition with other mutant cells (e.g. a single TET2-mutated clone in a “TET2-first” case). A larger JAK2 clone size may itself be more likely to give rise to a PV phenotype, but also is likely to increase the chances of JAK2 homozygosity arising. This association is demonstrated in our data, with a greater prevalence of 9pUPD in “JAK2-first” cases. However, phenotypic differences are seen even in cases without 9pUPD, suggesting other mechanisms still may be implicated.

These analyses still remain limited however, and do not allow us to fully explore the impact of clonal composition on disease phenotype, since they only fully determine mutation order in a subset of patients and do not explicate the whole clonal structure. As shown in Figure 18 there is potentially a great deal of complexity in even patients with 2 or 3 mutations. Ideally, this would be explored using single cell or single-derived colony approaches to determine (a) the true frequencies of each mutation order and the frequency of biclonality, (b) the impact of this on phenotype, and (c) the effect of differing clonal proportions (e.g. double vs single mutants).

5 Modelling of clinical outcomes in MPNs

5.1 Introduction

Broadly speaking, the aims of creating prognostic models are two-fold. Firstly, most modelling methods will provide an estimate for the effect an individual variable has on a patient's risk of developing a particular outcome (death, disease progression or treatment failure, for example). This can provide important information regarding causality and therefore inform a better understanding of biological mechanisms underpinning disease processes. In the context of MPNs, two important outcomes are transformation of chronic phase disease to myelofibrosis, and AML transformation. While a number of studies have examined paired chronic phase and post-leukaemic transformation samples, and have examined the risk factors for AML transformation from MF, few have comprehensively examined genomic factors present pre-transformation in patients with ET or PV that may have a causal role in subsequent transformation to MF or AML.

Secondly, prognostic models are clinically useful in that they allow us to estimate an individual's risk which may then inform treatment decisions. For example, patients with MF are typically risk-stratified using the dynamic international prognostic scoring system (DIPSS)[236] and the decision whether to treat conservatively or proceed to stem cell transplantation may hinge on this risk score. An important distinction here however is that variables may be present that provide useful predictive power without actually indicating a causal association – for example, anaemia is shown to be a predictor of poor overall or progression-free survival, but it is likely that this is because it is acting as a surrogate marker, for example for the degree of marrow failure or dysplasia. In this chapter, we examine a number of models and modelling methods in an attempt to both identify a set of variables that are likely to have a causative role in MPN disease progression, but also to build models that can best predict an individual patient's risk.

5.2 Evaluation of previously reported prognostic scoring systems

1875 patients had clinical follow-up (median 7.8 years, range 1 month to 43.6 years, total >16000 patient years) and therefore the impact of genomic variables on overall survival, myelofibrosis- and AML-progression free survival could be assessed for these patients.

For patients with myelofibrosis, a number of scoring systems have been developed to identify poor risk patients. The most commonly used of these are the international

prognostic scoring system (IPSS)[237], dynamic IPSS (DIPSS)[236] and DIPPS-plus[238]. More recently the Mutation-Enhanced International Prognostic Score System (MIPSS) has been developed[239]. Of these, only the latter two encorporate genomic variables, specifically the presence/absence of adverse cytogenetics and CALR and "high molecular risk" (HMR) mutations[234]. Unfortunately, it was not possible to compute these scores for the majority of the patients in this cohort, as data regarding blast percentage and constitutional symptoms were not available for the majority. IPSS and DIPSS were available for patients with MF from an external validation cohort, which is discussed later, allowing comparisons to still be made with these scoring systems. Two purely genomic classifications for patients with myelofibrosis have also been defined, one using only the CALR and ASXL1 mutation status[240], the other using the number of HMR mutations (ASXL1, EZH2, IDH1/2, SRSF2), each dividing patients into 3 risk groups[234]. More recently, the GIPSS (genetically inspired international scoring system) has also been proposed which includes CALR mutation type, ASXL1, SRSF2 and U2AF1 mutations and very unfavourable or unfavourable karyotypes, but this score was not addressed in this study[241].

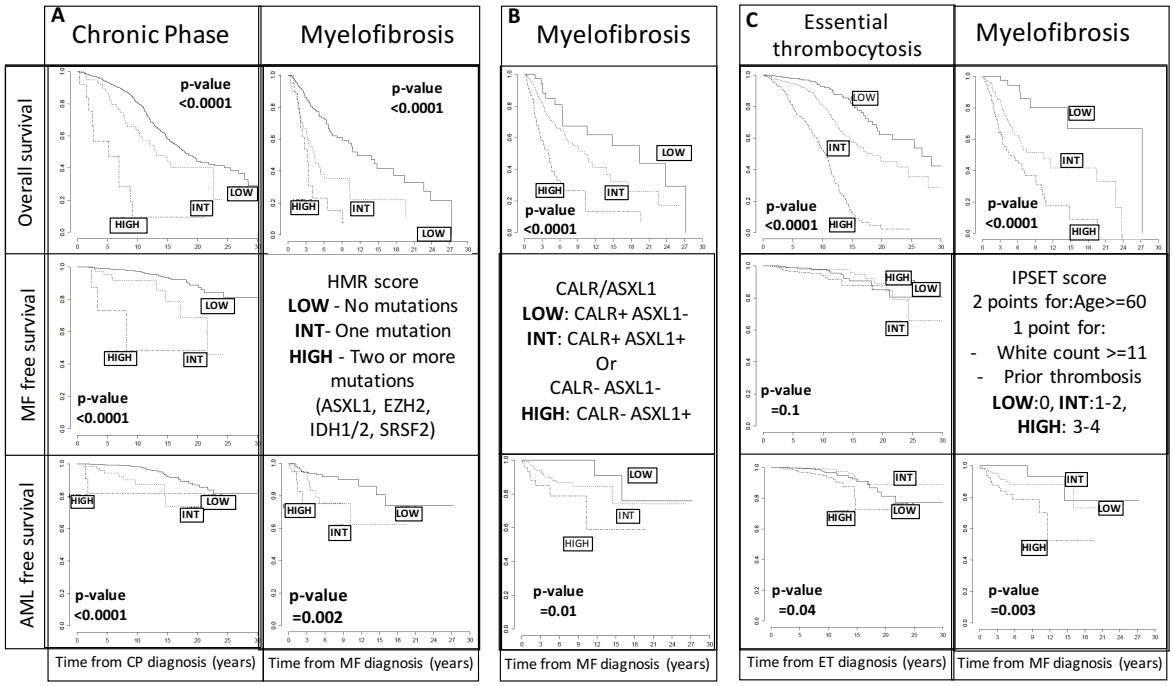


Figure 22: *Stratification of patients into high, intermediate or low risk groups according to (A) high molecular risk, (B) CALR/ASXL1 and (C) IPSET scores, as defined in the figure. Outcomes (overall survival, AML transformation and myelofibrosis transformation, for CP patients) are shown according to risk group.*

Figure 22 shows the survival curves for patients according to these two scores for patients with MF, demonstrating significant differences across these three groups using both HMR

or CALR/ASXL1 for stratification for both overall survival and AML-progression free survival. Additionally, HMR stratification was a significant predictor of overall, AML- as well as MF-progression free survival for chronic phase patients, while CALR/ASXL1 stratification was not informative (data not shown).

There are few prognostic systems proposed for chronic phase, with most classifications geared towards prediction of thrombotic risk. The international prognostic score for ET (IPSET) has been proposed, and also stratifies patients into three risk groups, based on age, white cell count and history of prior thrombosis[242]. When used to stratify patients with ET in this cohort, significant differences in overall survival are seen, but IPSET poorly correlated with risk of MF or AML transformation. IPSET additionally significantly correlated with overall survival, as well as AML transformation risk in patients with MF, in keeping with the observation that age and white count are predictors of outcome in MF, and are sufficient to provide reasonable discrimination.

These analyses serve to validate a number of pre-existing prognostic scores and suggest outcomes in this cohort are not strikingly dissimilar to those in other studies. The discriminatory ability for the HMR score in chronic phase patients, in addition to those with myelofibrosis, is an indicator that genomic variables may be of value in these patients as well, and may play a role in predicting not only survival but also disease transformation risk.

However, while these existing models have clear utility and are both easy to understand and implement they have a number of limitations (discussed below). In order to address these issues as much as possible, we retained a large number of variables, without dichotomisation of continuous variable (where possible), and fitted models for each disease state transition.

5.3 Assessment of relevance of clonal versus sub-clonal mutations

Since genomic variables are incorporated into our genomic classification as either ‘present’ or ‘absent’ regardless of clone size, patients may be classified into a TP53 group, for example, with only a small TP53 clone. This may be problematic when exploring the prognostic impact of these mutations, since a small clone is treated the same as a large one despite potentially having different clinical consequences

To assess the importance of clone size for clinical outcomes (and therefore whether to include this information in the multistate models), we modelled the impact on survival and

disease progression of mutation clone size (measured as a proportion of total tumour clone size in the sample), taking into account age, sex and MPN subtype. Of ASXL1, TET2, EZH2, CBL, SRSF2 and TP53, significant differences were only seen with ASXL1, with differences in overall survival (p-value 0.003), and AML transformation risk (p-value 0.01).

This observation that would require further validation given the relatively weak strength of association and the multiple hypothesis testing.

5.4 Univariate analyses

Across the MPNs different disease transitions are possible and we sought to model each separately. These were as follows:

- Death from chronic phase disease (with censoring at transformation to MF/AML)
- Death from myelofibrotic phase disease (with censoring at transformation to AML)
- Transformation from chronic phase disease to myelofibrosis (with censoring at transformation to AML or death from another cause)
- Transformation from chronic phase disease to AML (with censoring at transformation to MF or death from another cause)
- Transformation from myelofibrotic phase disease to AML (with censoring at death from another cause)

ET/PV/MF status was used as a time dependent variable, which could therefore change if a patient transformed to MF, which also allowed these transformed patients to be encorporated in the models used for MF overall survival or AML transformation risk. As an initial overview, table 8 shows the hazards ratios (restricted to those >2) for genomic variables significantly (alpha=0.05) associated with each outcome in univariate analysis.

	OS (CP)	CP to MF	AML (CP)	OS (MF)	AML (MF)
NRAS	111.0			10.4	6.5
SRSF2	8.6	22.2	34.9	2.3	
CBL	2.5	6.2	5.6	4.8	4.3
IDH2	2.2	4.8	5.5		
ASXL1	2.1	3.0	3.4	2.6	3.0
TET2	2.0	2.7	3.8		
TP53			9.3	2.3	6.2
GNAS			6.8	4.6	9.7
GNB1	7.0				
MBD1	6.7		22.2		
U2AF1	4.1		14.7		
CUX1		16.6			
ZRSR2		15.0			
SH2B3		10.1			
KRAS		9.0			
PHF6		8.9	8.5		
EZH2		4.1		3.0	
NFE2		3.2			
MPL		2.7			
RUNX1			7.9		
PPM1D			2.6		
DNMT3A			2.6		
1p	2.4				3.1
17p			6.2	2.9	8.5
5q			5.8	4.5	15.5
14q		23.1			
7q			7.0		
Trisomy 9			6.9		
11q				7.3	
4q				4.4	

Table 8: *Hazards ratios for significant covariates in univariate analyses for overall survival in CP, CP to MF transformation. CP to AML transformation, overall survival in MF and MF to AML* transformation

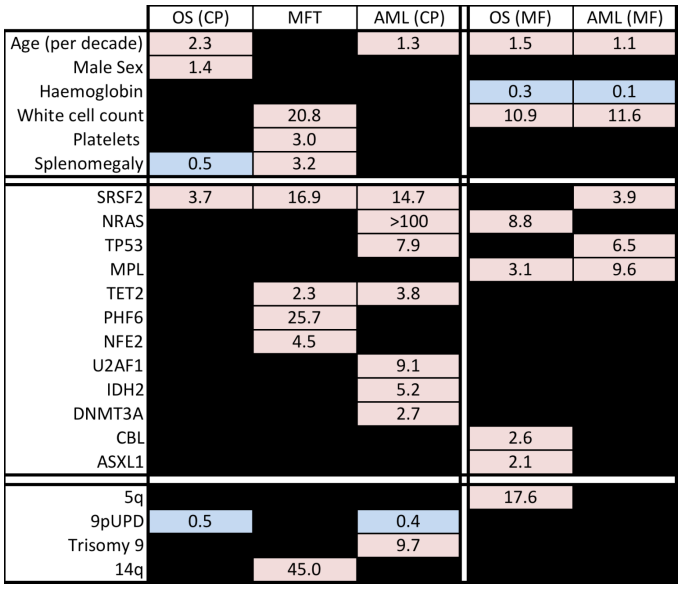
This superficial analysis suggests a number of new candidate prognostic markers. In particular, NRAS, CBL and TET2 appear to be strongly associated with poor outcomes generally, and TP53 with AML transformation. Furthermore it suggests that different variables may be important in different disease states or for different disease transitions.

5.5 Stepwise variable selection

Proportionality was tested for each continuous variable and there was no evidence of violation of the proportionality assumption. Therefore Cox proportional hazards modelling was used to fit models for each disease status transition.

One danger of this approach, given the large number of variables, is of over-fitting. One suggested approach is that of variable selection, which has the further utility of simplifying the model and making it more parsimonious and easier to implement in clinical practice. This was performed using backward variable selection using the Bayesian Information Criterion (BIC)[243] at each variable selection step as the criterion for whether removal of a variable improved/worsened the model. The minimal model in each case was required to include age, to avoid confounding for age. Table 9 shows the

variables selected for each model.



	OS (CP)	MFT	AML (CP)	OS (MF)	AML (MF)
Age (per decade)	2.3		1.3	1.5	1.1
Male Sex	1.4				
Haemoglobin				0.3	0.1
White cell count		20.8		10.9	11.6
Platelets		3.0			
Splenomegaly	0.5	3.2			
SRSF2	3.7	16.9	14.7		3.9
NRAS			>100	8.8	
TP53			7.9		6.5
MPL				3.1	9.6
TET2		2.3	3.8		
PHF6		25.7			
NFE2		4.5			
U2AF1			9.1		
IDH2			5.2		
DNMT3A			2.7		
CBL				2.6	
ASXL1				2.1	
5q				17.6	
9pUPD	0.5		0.4		
Trisomy 9			9.7		
14q		45.0			

Table 9: *Variables retained in variant selection modelling using Bayesian information criterion at each selection step for outcomes shown in Table 8.*

These data support poor risk associated with increasing age, leucocytosis and SRSF2, as well as NRAS and TP53, regardless of whether the patient is in chronic phase or not. Thrombocytosis, leucocytosis, splenomegaly, TET2, NFE2 and PHF6 mutations and 14qUPD are additionally identified as candidate predictors for myelofibrotic transformation.

Even restricting risk stratification to whether any of these high-risk mutations are present or not provides good discrimination for both MF and AML transformation from chronic phase with HRs of 5.0 (95% CI: 3.0-8.2) and 5.7 (95% CI: 4.7-8.7) respectively, confirming that genomic-based risk stratification outside of the context of myelofibrosis is possible.

5.6 Random effects Cox proportional hazards modelling and variance contributions

Variable selection however, is thought by many to be a flawed statistical method since it results in loss of information available in a fully-fitted model and to an over-estimation of effect sizes and under-estimation of p-values, among other reasons[244].

To better fit a model and reduce over-fitting, an alternative to variable selection is using "shrinkage" methods, which impose restrictions on effect sizes to ensure they are not over-estimated. Methods include lasso and ridge regression. The method used in this analysis,

using random effects, is similar to a ridge regression and was developed by Dr Moritz Gerstung. Implementation was carried out by Drs Moritz Gerstung, Rob Cantrill and Jacob Grinfeld.

The methodology of random effects Cox proportional hazards models has previously been reported in the context of AML[215]. In brief, these hierarchical models require the partitioning of co-variates into groups (in this case: genetic, cytogenetic, clinical, demographic and nuisance/cohort), and assume a normal distribution for the parameters in each group. This results in a ridge-type regularisation reducing the bias that would result from using a large number of covariates.

Using ET/PV/MF as time-dependent variables, and after standardisation of continuous variables (so that coefficients can be more directly comparable), individual fits were again created for each disease transition. Cross-validation (described later) suggested over-fitting particularly for the MF to AML fit (only 31 transformation events), and therefore a combined model for both CP and MF patients was ultimately used (86 transformation events). Alternative models were also made that included gene-gene interactions, and dichotomisation of JAK2, CALR, ASXL1, TET2 and DNMT3A into high and low allele fractions, but these did not show significant improvement in prognostic accuracy.

Figure 23 shows the four disease transitions modelled and the contributions of each variable to the overall variance of the calculated risk for each model. This calculation accounts not only for the model co-efficient (hazards ratio) for a given mutation, but also for the frequency of that mutation. This is particularly useful for the creation of a prognostic system, as there maybe strongly significant variables, only found very rarely in a cohort, and conversely other variables only associated with a small increased risk may still be important to test for if they occur commonly (TET2 and DNMT3A may be examples of this).

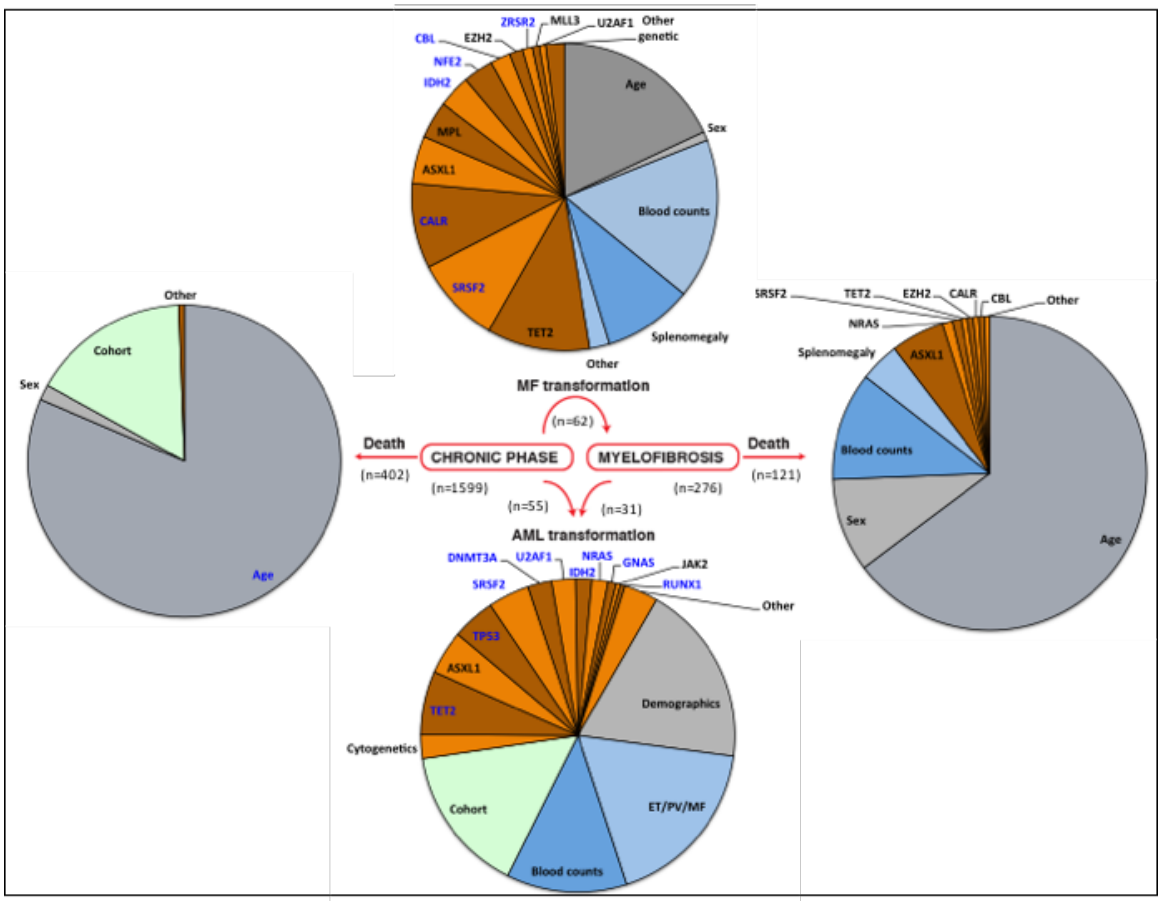


Figure 23: *Individual fits for disease state transitions. Pie charts show variance contributions for each variable. Variables in blue are associated with hazards ratios >2.*

This analysis suggests that genomic variables have a minimal contribution to predicting non-progression related mortality risk in myelofibrosis once age, sex and blood counts are taken into account (although ASXL1, NRAS, TET2, CALR mutations may contribute), and that they have even less of a role for non-progression mortality prediction in chronic phase. However, genomic variables strongly contribute to AML transformation risk for CP and MF phase patients and for MF transformation from CP. Confirming the results of previous analyses from this cohort and previously proposed prognostic models, ASXL1, EZH2, IDH2, SRSF2 and U2AF1 are associated with poor prognosis/disease transformation. Additionally, prognostic roles for NRAS, TP53, TET2, DNMT3A, CBL, NFE2, RUNX1 and GNAS are identified and the presence of either CALR or MPL was shown to be predictive for MF transformation in ET (HR 2.4, 95% CI 1.4- 4.3, p-value=0.002).

5.7 Generation of multi-state model

In isolation, these models can therefore provide information on the contribution of each

variable to the risk of each disease state transition. In clinical practice however, it would not be practicable to calculate 2 or 3 individual predictions for each patient and it is not evident how these predictions should be integrated. We therefore derived an integrated multi-state prediction model using the schema shown in Figure 23. At each time point (selected as roughly every 15 days, $1/25^{th}$ of a year), a patient's risk of transitioning into another disease state is calculated and this process is carried out iteratively over a 25 year period. Patients are either defined either as (a) starting in chronic phase, and therefore at each point have a certain probability (defined according to their individual risk calculated from the individual random effects Cox proportional hazards models) of transitioning to death or AML (terminal states) or to MF, from where they can also transition to death of post-MF AML or (b) start directly in MF and can transition to death or AML. AML itself was a terminal state - in other words the risk of further transition from AML to death was not calculated, both because of the limited data available to do so, but also because this was not felt to be essential from a clinical perspective. (This algorithm was developed by Dr Moritz Gerstung with implementation by Dr Rob Cantrill, and further modification and downstream data processing by Dr Jacob Grinfeld).

This algorithm thus generates a 366x5 (for patients starting in CP) or 366x2 matrix (for MF) for each patient, where 366 corresponds to $1/25^{th}$ year "snapshots" over a 25 year period from diagnosis together the probability of being in the following states: Death from CP, Death from Post CP-MF, Death from Post-CP AML, Death from AML from post-CP MF, Alive in Post-CP MF (for CP) or Death from MF (as starting state) or Death from post-MF AML (for MF). The probability of being alive without transformation at an individual timepoint is therefore calculated as 1 minus the sum of these (5, or 2, depending on starting diagnosis) probabilities. Predictions generated are presented using leave-one out cross-validation (i.e. for each patient, the model was built on the remaining patients to avoid over-fitting).

Examples of predictions generated in this manner are shown in Figure 24. Figure 25 shows predictions of event free survival for equally-sized subsets of patients with either CP or MF, compared to the overall outcomes of these groups. This method has a number of advantages – it integrates the information from the 4 individual predictive models into one set of predictions; it provides patient specific predictions using the full dataset (without dichotomisation of continuous variables); and it allows predictions to be obtained for individual outcomes for specific time-points.

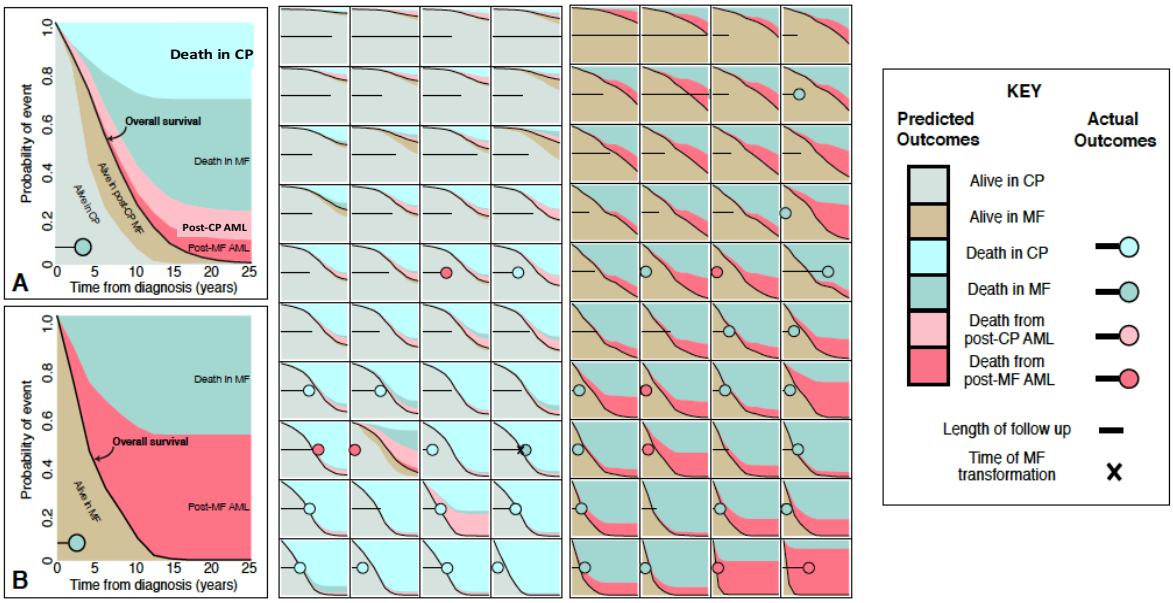


Figure 24: Personalised predictions of patient outcome. Each of the tiles represents the personalised predicted outcome of an individual patient. Two tiles **(A)** and **(B)** have been enlarged for illustrative purposes. The top left panel (A) depicts the predicted outcomes of a 79 year old female patient who presented with ET with haemoglobin 104g/l, white cell count $8.4x10^9$/l and platelet count of $2300x10^9$/l, mutations in *CALR*, *SRSF2*, *IDH2* and 18q LOH. The varying probabilities of each of these transitions can be judged from the vertical axis and their respective Kaplan-Meiers over a 25 year time period shown along the horizontal axis. The labelled black curve shows the predicted Kaplan-Meier curve of overall survival. The patient in (A) transformed to myelofibrosis (MF) and died within 5 years, as indicated by the line and circle in the bottom left. Panel B shows the predicted and actual outcome of a 57 year old male patient diagnosed with MF with Hb 125g/l, WCC $27x10^9$/l and Plt 119 $x10^9$/l, mutated *TET2*, *ASXL1*, *CBL* and *BCOR* along with chr7q- and 11q-. This patient died in MF within 2 years. The remaining tiles show the predictions and outcomes for 40 patients in CP and 40 patients in MF selected at equal intervals when ranked by increasing risk for progression or death (top left to bottom right for each set).

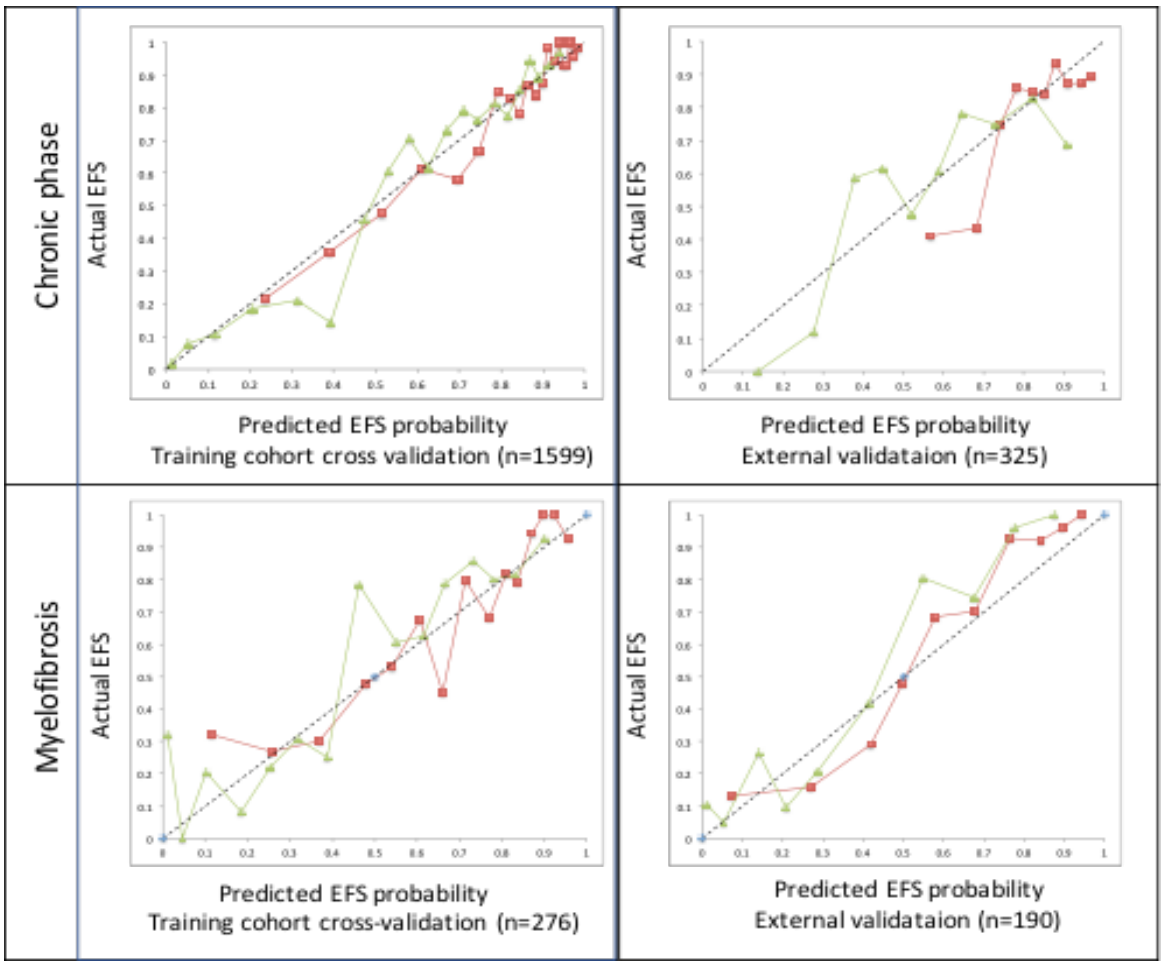


Figure 25: *Predictions for equally sized subsets of each cohort, ranked according to their predicted risk according to the multi-stage model. Predicted event free survival (x-axis) is plotted against observed outcomes for each group as a whole (y-axis), for chronic phase (top two panels) and myelofibrosis patients (bottom two panels). These predictions generally follow the dotted line (where predictions are equivalent to actual outcomes) even in external validation cohorts with limited genomic data. 10 year predictions (brown) and 15 year predictions (red) are shown.*

Both Figures 24 and 25 demonstrate that this modelling method can generate continuous, accurate predictions. An additional advantage of this approach is that, since the predictions are continuous variables and can be provided for any time-point, the use of the model, in theory, can be tailored to the clinical need (e.g. maximisation of specificity or of sensitivity, stratification into an number of risk groups, identifying the x% of highest/lowest risk patients etc).

Figures 26 and 27 show examples of stratification into two groups using cut-off points (for 10 year predictions) that maximise sensitivity and specificity using a Receiver-operator characteristic analysis.

A more comprehensive evaluation of model performance relative to other proposed models is provided in **Section 5.10**.

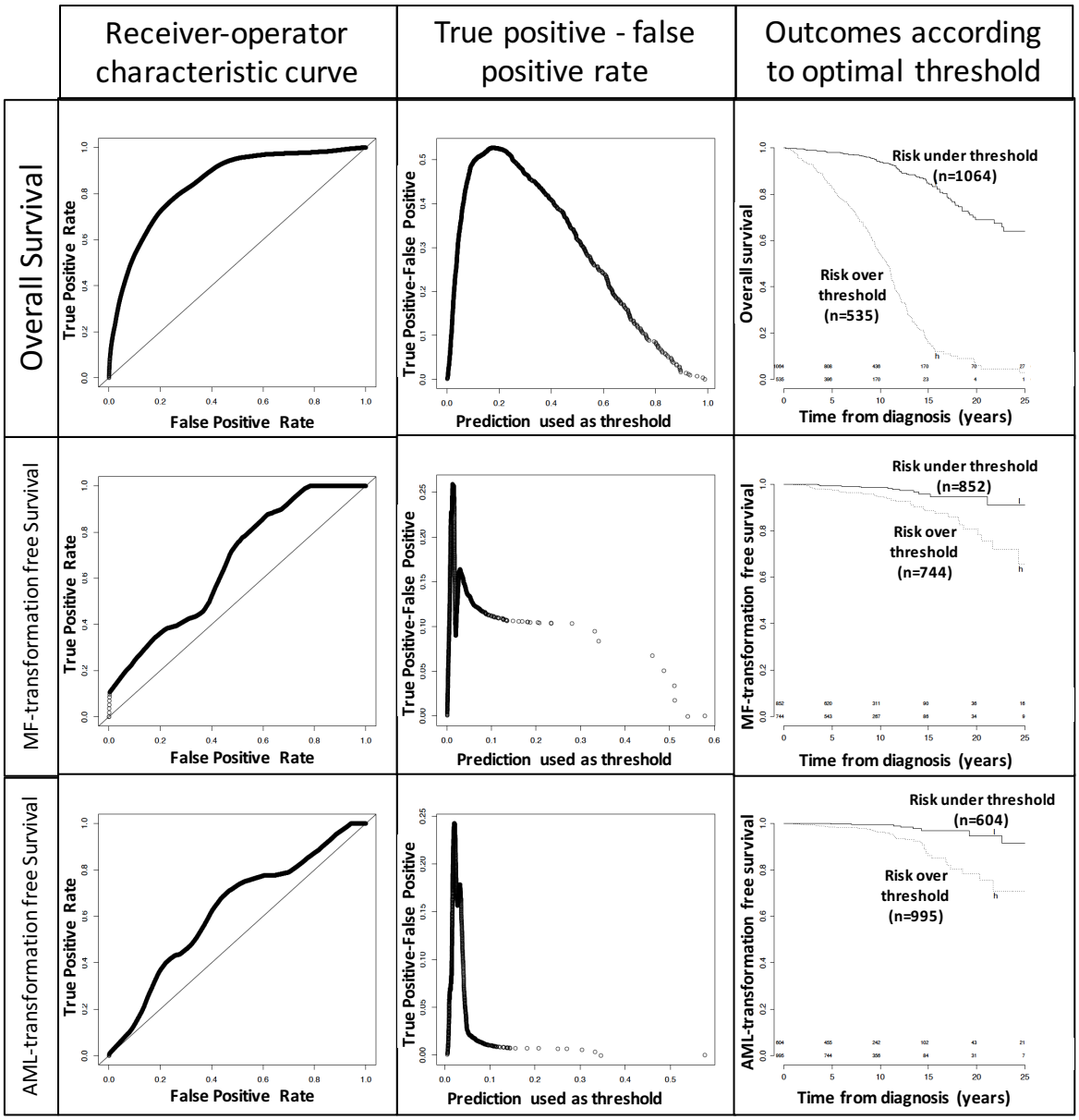


Figure 26: *Predictions at 10 years for overall survival, MF transformation and AML transformation in chronic phase are used in each case to generate (left) receiver operator characteristic curves, (middle) curves of true positives-false positives in order to find cut-offs maximising sensitivity and specificity, and (right) Kaplan-Meier curves using this cut-off to divide patients into high and low risk groups.*

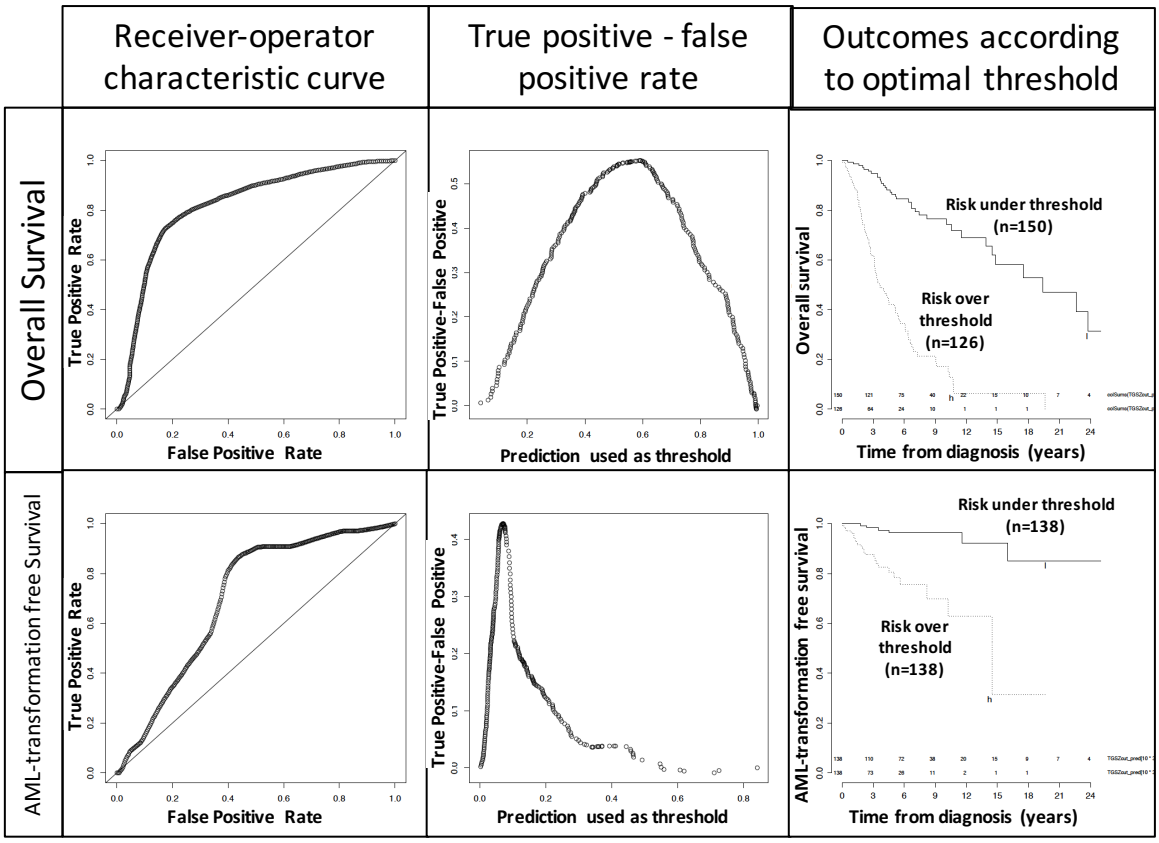


Figure 27: *Predictions at 10 years for overall survival and AML transformation in myelofibrosis are used in each case to generate (left) receiver operator characteristic curves, (middle) curves of true positives-false positives in order to find cut-offs maximising sensitivity and specificity, and (right) Kaplan-Meier curves using this cut-off to divide patients into high and low risk groups.*

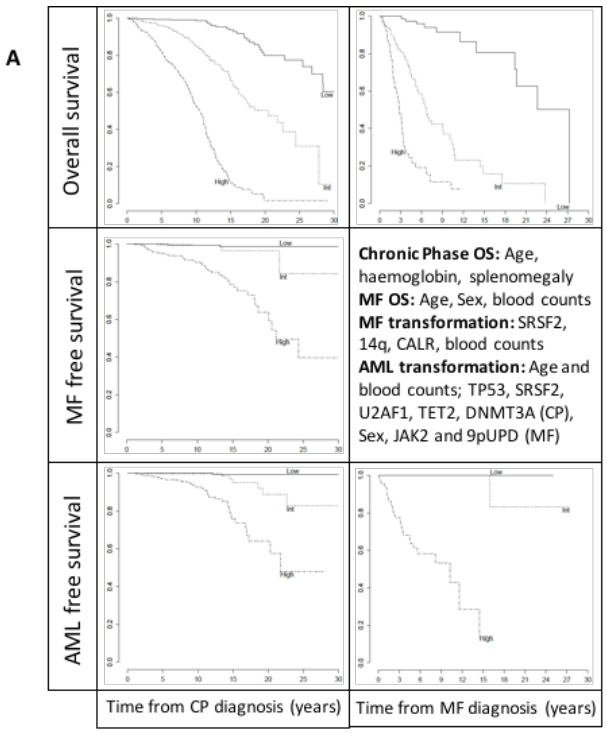
5.8 Alternative modelling prognostic classifications

5.8.1 Decision trees / Recursive partitioning

In order to validate these results and generate a potentially superior predictive model, a number of alternative modelling approaches were also attempted. Decision tree analyses recursively partition the predictor variables to create a branching tree structure, attempting to determine which variable (and its cut point for non-binary variables) will give the most informative split between the patients categorised into each of the resulting branches (defined using statistical comparisons between the two branches or measures of impurity/dissimilarity). Once the algorithm has determined the optimal decision tree, this effectively provides a flow diagram, with decisions at each node, allowing each patient to be classified into different risk groups.

These approaches have a number of advantages: they are non-parametric approaches that

do not depend on the distributions of continuous variables (and therefore transformations/splines are not required), offer a method for dichotomisation of variables, are easy to interpret and apply by just using the resulting flow diagram and inherently allow for interactions between variables to be shown (e.g. a variable may only be predictive in a certain subset of patients, i.e. down one branch). As an example of the latter point, the predictive model for AML transformation from CP classifies the highest risk groups as TP53/SRSF2/U2AF1-mutated patients over 48, or patients over 72, suggesting (providing this model is correct and validates) testing for these mutations is not helpful in the under the age of 48 (where haemoglobin is a better predictor) or over the age of 72 (since further splitting did not improve the model).



B

	OS (CP)	MFT	AML (CP)	OS (MF)	AML (MF)
White cells	3	3	10	2	5
Haemoglobin	4	4	8	4	4
ASXL1	9	6	12	6	3
TP53	14	18	2	11	1
JAK2	7	13	9	18	14
Age	1		5	1	2
SRSF2	6	1	1	10	
U2AF1	15		4	17	10
9pUPD	10	10	18	9	
Platelets	5	7		5	
TET2	8	8	3		
ET/PV	2	11	11		
7q			13	12	8
5q			15	13	6
GNAS			14	15	9
NRAS				3	12
IDH2		12	6		
EZH2		14		7	
CBL	12			8	
NFE2		5			18
MPL	13	9			
1pUPD			17		7
14q	20	2			
Trisomy 9			7		21

Figure 28: *Results of decision tree and random forest analyses. (A) Kaplan Meier curves classifying patients into 3 groups according to predicted risk from rpart analysis of the training cohort. (B) The rankings of the top ranked variables according to Random forest variable importance analysis.*

From Figure 28A it can be seen that these models are very successful when re-applied to the training data. However, they are very prone to over-fitting, and are greedy algorithms, so that early splitting may lead to suboptimal solutions. Part of the success of these models on the training data is that disparate, seemingly arbitrary groups can be ranked together as being high or low risk, e.g. for OS in MF, female patients under 62 without 9pUPD but with white counts>6.7, form a higher risk node, as do patients over 62 with platelet counts

over 190 and under 250, which strongly suggests a model that is over-fitted and clinically implausible. In most cases therefore, "pruning" of the trees and cross-validation are required. However, these models still may be helpful in terms of highlighting a list of important predictive variables. As can be seen from Figure 28, the variables selected do overlap with many highlighted by other analyses.

The *ctree* algorithm (which uses a different splitting criterion) was also used but gives far fewer branches. However, these analyses also highlight a prognostic role for age, sex, haemoglobin, white cell count, platelet count, TP53, SRSF2, U2AF1 and 1pUPD.

5.8.2 Random forests

The use of random forests avoids the main pitfalls of decision tree algorithms. Random forests are created by "growing" a large number (e.g. 500) of individual decision trees, each time using a random subset of patients and a random subset of the available variables. The random selection of variables and patients for each independent tree reduces the chance of over-fitting the model, and the fact that for each iteration there are "out of bag" patients that the model can be tested on, means that cross-validation and estimation of the model's error can automatically performed along with the creation of the model.

The performance of random forest models is discussed in section 5.9, but an additional benefit of this process is that the importance of each variable can also be computed by this process, essentially by assessing by how much the model's error changes when the variable is not used. Figure 28B shows the rank for each variable according to this analysis, for each outcome assessed, again emphasising the prognostic role for age, sex, blood counts, and mutations in epigenetic regulators, spliceosomal and cytokine signal transduction components, NRAS and TP53.

5.9 Evaluation of model performance and model comparisons

A number of measures were used to evaluate and compare the prognostic models discussed above, including measures of concordance and Brier scores. These are discussed in detail in **Section 2.8.6**.

The aim of these analyses is to compare (a) the different methods for modelling clinical outcomes (e.g. the full multistage model using random effects Cox proportional hazards modelling relative to simplified models using variable selection methods), (b) the effect of

different variables (e.g. to determine whether an extended genomic characterisation improves accuracy of prediction relative to simpler models) and (c) to compare the models proposed in this study with existing prognostic models, such as the HMR score.

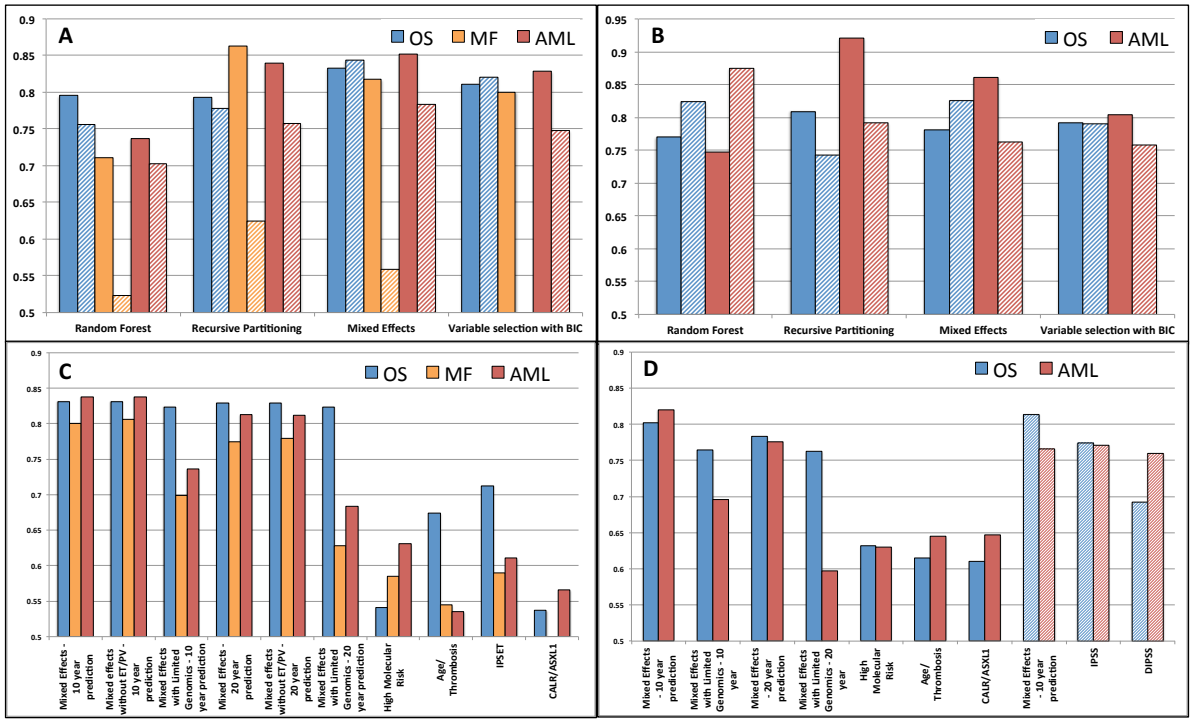


Figure 29: *Concordances (Harrell's c-statistic) for different models assessed. (A) and (B) show comparisons between individual Random effects Cox proportional hazards models against those derived using random forest, recursive partitioning and variable selection using BIC, for each outcome assessed for CP and MF patients respectively. Paler bars represent the results of applying these models to the external dataset. Figures (C, chronic phase) and (D, myelofibrosis) show comparisons of the full cross-validated multi-stage model (leave-one-out approach) with equivalent models that limited the genomic variables to JAK2, CALR or MPL, or removed the distinction between ET and PV, as well as HMR, IPSET, models with age and prior thrombosis or CALR and ASXL1, IPSS and DIPSS.*

Comparisons of concordance are shown in Figure 29 and Appendix 4. A number of conclusions can be drawn from these

i) Random effects Cox proportional hazards modelling performs better overall (in terms of concordance and degree of over-fitting for individual models) than other methods used, including Cox proportional hazards models with variable reduction, decision trees and random forests, although in some cases these

differences are negligible. Random forests also perform well, however.

ii) The inclusion of extended genomic characterisation improves the performance of multi-stage Cox proportional hazards models in the prediction of MF or AML transformation from CP, AML transformation from MF and, to a lesser extent of OS in MF.

iii) The distinction between ET and PV in CP patients does not appear to significantly affect the performance of multi-stage models when other variables (including blood counts, age and splenomegaly) are taken into account.

iv) Predictions for overall survival in CP and MF tend to be more accurate than predictions of AML or MF transformation.

v) The time-point chosen for multi-stage predictions in order to rank/classify patients does not appear to be important as the ranking appears to stay relatively constant. However, the accuracy of predictions appears to be greatest at earlier time points, which is likely to relate to the fact that the uncertainty inherent to the prediction is lower then (i.e. fewer events are expected to happen).

vi) Predictions derived from Mixed effects multi-stage modelling show better risk stratification than the HMR score, and models that use age>60 and thrombosis alone, CALR and ASXL1 alone, IPSET, IPSS and DIPSS. However, comparisons with IPSS and DIPSS could only be made in the external validation cohort, which was not only smaller, but had more limited genomic characterisation.

It is reasonable to assume that the performance of the multi-stage model is actual under-estimated here (and therefore that the improvement relative to DIPSS/IPSS would be greater) if full genomic information was available, and this appears to be the case when performance in this cohort is compared to results of cross-validation in the training cohort.

Regarding point (vi) it is worth noting that the multi-stage model provides continuous predictions and therefore it is rare to have predictions tied across different patients. However this is not the case for scores such as IPSS/DIPSS for which comparisons can only be made for patients across different risk groups. While this is a potential benefit for

the multi-stage model, this may result in a flawed comparison between groups. Therefore, direct comparisons were made between HMR, IPSS and DIPSS and discrete groups defined by ranking patients by their multi-stage predictions and splitting them in the same proportions as HMR/IPSS/DIPSS. The results of this are as follows:

- Multi-state (stratified into 3 groups, n=158/75/45) compared with HMR (3 groups) assigns 65% of patients to the same risk group (with 17% risk group increased, 18% risk group decreased). Concordance for overall survival improves from **67.1** to **78.2%.** Concordance for AML-transformation improves from **66.6** to **76.6%.**

- Multi-state (stratified into 4 groups, n= 85/94/63/56) compared with IPSS (4 groups) assigns 58% of patients to the same risk group (with 22% risk group increased, 20% risk group decreased). Concordance for overall survival improves from **77.4** to **80.0%**. Concordance for AML-transformation improves marginally from **77.1** to **78.5%**.

- Multi-state (stratified into 4 groups, n= 43/89/109/57) compared with DIPSS (4 groups) assigns only 44% to concordant risk groups (with 30% risk group increased, 27% risk group decreased). Concordance for overall survival improves from **69.2** to **78.1**%. Concordance for AML-transformation worsens very marginally from **76.0** to **75.9%**.

This confirms the results of the earlier analysis showing that personalised multi-stage predictions provide, in almost all cases, superior risk stratification than the existing prognostic models assessed.

5.10 Implementation of individualised patient prediction calculator

So that the multi-stage predictions could be available as an alternative or adjunct to current scoring systems and its use could be trialled prospectively, an implementation of the model was made using the R package shiny, which can be run in a browser at https://cancer.sanger.ac.uk/mpn-multistage.

This package utilises two files, server.R, which determines how user-inputted data is processed, and ui.R, which determines how the input (and which is then passed to server.R) and output (where the results from server.R are displayed) are displayed and associated with variables in server.R. In this case, users can input patient-specific data (e.g. age, presence/absence of mutations), which are then entered into the multi-stage

model and probabilities for each time point are generated. From this, predicted survival curves are drawn and presented to the user, along with estimates for median event free survival, and probabilities for each outcome at 5, 10 and 20 years. Additionally the (anonymised) predictions for patients in the training set are available, together with their actual outcomes can be viewed.

5.11 Discussion and Future work

In this chapter we examined the variables that most strongly determine the risk of death, MF and AML transformation in MPNs. A variety of different modelling methods are used and compared. Common to the majority of the prognostic systems examined is the confirmed predictive role for Age, sex, blood counts, chromatin-regulator/spliceosome mutations (ASXL1, SRSF2, U2AF1, EZH2), NRAS, CBL and TP53. Although carrying a substantial increase in risk, these mutations are relatively uncommon, particularly in chronic phase.

There is clearly significant overlap between mutations found in the chromatin/spliceosome group (defined purely using the genomic data alone), mutations associated with an MF phenotype at disease presentation (or time of earliest available sample where not available), and the mutations shown here to predict MF transformation from CP, AML transformation from CP or MF, and, to a lesser extent death in MF. As discussed previously, the mechanisms by which these mutations may lead to disease progression need to be explored further, but may include acceleration of the myeloproliferative phenotype itself, or alteration in the underlying phenotype (for example through alteration of the balance between differentiation and stem cell expansion, or through differential secretion of pro-inflammatory/pro-angiogenic cytokines).

The results of the analysis here additionally suggest predictive roles for more common variables which generally are not restricted to a particular MPN subtype, including 9pUPD, MPL, CALR, TET2 and DNMT3A. For these variables, the associated increase in risk may be lower, but their frequency means that they still have an important effect population-wide, and are therefore likely to be useful to still include in a model.

Analysis of the effect of clone size on risk suggests that for most genes, whether a mutation in a given gene is clonal or subclonal does not substantially affect its impact on clinical outcome. This suggests that in the MPNs, the major determinant of clinical outcome is likely to be affected more by most aggressive subclone a patient has, rather

than specifically the genomic changes of the dominant clone. One exception in this analysis was ASXL1, however this observation would require further validation given the relatively weak strength of association and the multiple hypothesis testing.

Although the multi-stage model performs well, and has a number of advantages over existing scoring systems, there are a number of further steps that should or could be taken before it could be used for clinical practice. Firstly, the risk of over-fitting is always a concern. Even though internal cross-validation was used, and validation was carried out on an external cohort, the external cohort had only limited genomic characterisation. Therefore it is important to validate this model in other cohorts of patients. However, the model itself might be improved in a number of ways. Beyond validation, data from a wider pool of patients could be used to expand the knowledge bank from which the models are built, thereby refining the co-efficients used and improving the power to detect smaller effects. Secondly, the variables used could be expanded to include factors such as other genomic variables (e.g. chromosomal translocations or the clone sizes for detected mutations), degree of bone marrow fibrosis or other histopathological variables, blood count parameters (e.g. blast count, neutrophil count or red cell distribution width[245]), germ-line variants (e.g. JAK2 haplotype or rs11104870) or co-morbidities (e.g. cardiovascular risk factors). Finally, the modelling process might be improved, for example by including selected interaction terms or through transformation or the use of splines for selected variables[244]. For example, the effect of platelet count is modelled linearly, but one might expect poorer outcomes to be associated with very high or very low platelet counts.

6 Role of inflammatory cytokines in MPNs

6.1 Introduction

Haematopoietic stem and progenitor properties such as quiescence, self-renewal and differentiation are tightly regulated by micro-environmental factors. There is increasing evidence that there is dysregulation of these processes in myeloproliferative neoplasms (MPNs)[246].

As discussed in **Section 1.10**, stromal secretion of cytokines such as IL-6 or TNF-alpha may give a competitive advantage to the tumour which can further be supported through remodelling of the microenvironment induced by the tumour clone itself. Secretion of soluble mediators by the clone itself, or induced by the clone, may also drive fibrosis and angiogenesis as well as promoting DNA damage.

In general, across the MPN subtypes MF has been associated with the highest levels of IL-1RA, -2R, -6, -8, -10, -12,-15 TNFa, IP-10, HGF and VEGF. Of these, the levels of IP-10, IL-2R, -8, -12 and -15 are associated with an adverse prognosis[247], and IL-8 with increased microvascular density (a feature of MF)[248]. Overall therefore, it appears that pro-inflammatory signals may play a valuable role in MPN development and disease progression, and therefore a more comprehensive description of cytokine secretion across the MPN subtypes may offer insights into their biology as well as offer predictive biomarkers or therapeutic targets.

In order to investigate these possibilities, we quantified the levels of pro-inflammatory mediators in serum samples from over 400 patients with MPNs, as well as comparator samples from healthy controls.

6.2 Clinical and Genomic correlates of cytokine levels

6.2.1 Discovery 32-plex panel

Cytokine concentrations for 38 cytokines were assessed in an initial cohort of 199 patients and controls (100 patients with ET, 32 with PV, 53 with MF and 14 healthy controls) to identify those that were most discriminatory for MPN subtype.

Surprisingly, few cytokines were significantly different when concentrations were compared between control and MPN cases as a whole, after correction for age and sex: GRO-alpha (CXCL1) and EGF were most significant (p-values 0.01 and 0.03

respectively). However, IP-10, TNF-alpha, GRO, TGF-alpha, EGF, Eotaxin, IL-1RA, IL-8, IL-3 varied significantly across MPN subtypes ($p<0.005$ in all cases).

A discriminatory role for IP-10, TNF-alpha, GRO, TGF-alpha, EGF, Eotaxin, IL-1RA and IL-8 was confirmed upon multivariate analysis taking into account patient age and sex. Furthermore, these same factors contributed the most in random forest analyses predictive of diagnosis. IL-6 and IFN-γ were also taken forward for further analysis given previous evidence for their role in MPN pathogenesis, resulting in a total of 10 candidate cytokines subsequently measured using a 10-plex assays (Table 10).

	MPN and controls (accuracy)	MPNs only (accuracy)	MPN and controls (Gini)	MPNs only (Gini)	MPN vs Control	ET vs PV	ET vs MF	PV vs MF
IP.10	1	1	1	1		***	***	.
TGF.a	2	2	2	2		**	**	
GRO	3	4	3	4	*	**	*	
EGF	6	3	6	5	*		**	*
IL.1Ra	7	6	4	3				
Eotaxin	4	5	5	7			***	**
TNF.a	5	7	9	10			**	
IL.12.p40	12	10	7	6		*	*	
sCD40L	9	12	10	14			.	.
IL.8	8	8	27	9				
IFN.a2	15	9	22	8		**	**	
G.CSF	14	11	15	15		.	**	*
IL.6	11	13	18	17				
IL.10	21	16	17	12			.	
IL.3	19	26	8	19		.	***	
Fractalkine	16	19	20	20	.			
GM.CSF	20	22	24	28			*	
MDC	18	14	33	37			.	.
IL.9	28	29	30	16			*	.
IL.12.p70	29	33	16	32			**	
IFN.y	26	25	34	31				
IL.17	24	21	38	34			.	
FGF2	27	32	26	36			*	
IL.4	37	36	32	25		*	.	
Flt3L	35	34	37	35		.	**	

Table 10: *Rankings for variable importance in 2 random forest analyses classifying patients into diagnostic group (one with control patients, and one restricted to MPN patients alone. Ranking according to Gini and mean decrease in accuracy. Significance levels : .<0.1, *<0.05, **<0.001, ***<0.0001*

6.2.2 Validation of selected candidate cytokines

Results from the 10-plex assay were compared to those from the original 38-plex and found to be concordant (data not shown, r^2 values>0.9). Initial findings were then validated using an additional 110 patients (46 with ET, 19 with MF, and 41 with PV – a total of 305 patients across both cohorts). A subset of samples were repeated across plates, to allow correction for inter-plate variation, and additionally 25 patients had samples from 2 timepoints and 9 from 3 timepoints.

In order to account for the effect of inter-plate and inter-patient variation, sampling time relative to diagnosis, and the confounding effect of blood counts, age and sex, when identifying an association between cytokine levels and MPN subtype, mixed effects modelling was used. Here, diagnosis and sex were modelled as fixed effects but age, time from diagnosis, plate, and blood counts were modelled as random effects. The effect of diagnosis on a cytokine's levels was therefore assessed through comparison of models that did or did not include diagnosis as a variable, which allowed p-values for this comparison to be generated.

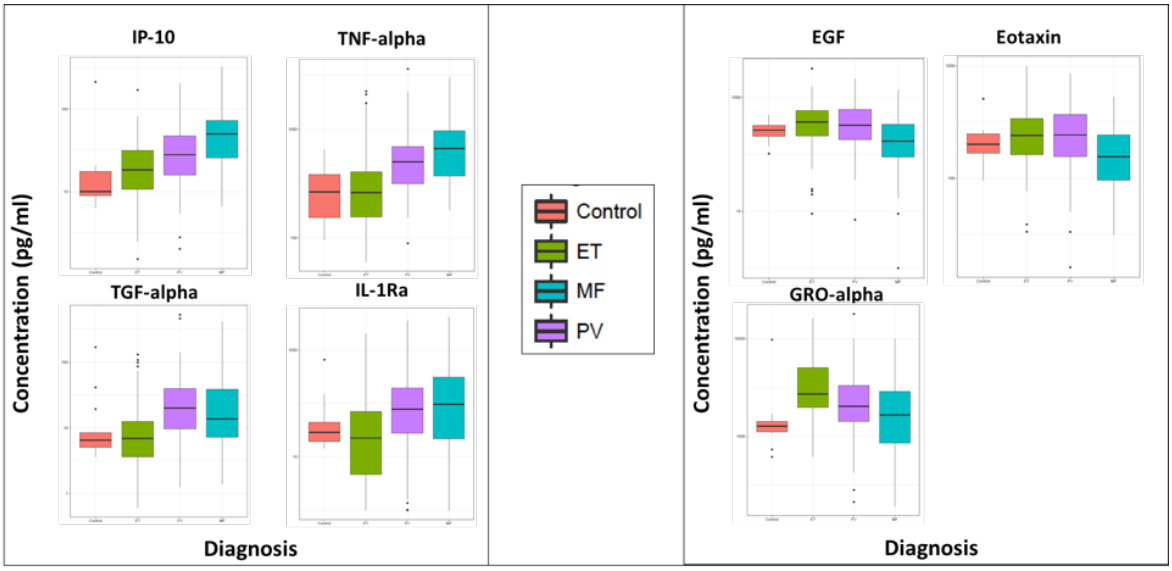


Figure 30: *Levels of IP-10, TNF-alpha, TGF-alpha, IL-1Ra, EGF, Eotaxin and GRO-alpha according to MPN subtype.*

Overall, the results of these analyses are in keeping with the concept that a continuum of increasing pro-inflammatory cytokine concentrations exists from ET to PV to MF:

Levels of TGF-a and IL-1Ra in patients with ET were lower than those in both PV and MF ($p<0.001$ in all cases) but did not show significant differences between PV and MF cases, Levels of IL-6 and IL-8 were significantly lower in ET than in MF (and did not differ

between PV and MF) in univariate analyses but did not meet the threshold for statistical significance in multivariate, mixed-effects analysis (data not shown). Levels of IP.10 and TNF-alpha showed a clear continuum, increasing significantly from ET to PV to MF (Figure 30).

In contrast, GRO-alpha levels were markedly raised in ET patients compared to PV, MF and control cases ($p<0.001$) while EGF and eotaxin were higher in chronic phase cases than MF cases ($p<0.001$).

6.2.3 Genomic associations

Samples from a final cohort of patients were analysed from the 10-plex assay to test an additional set of hypotheses regarding genomic associations and the predictive value of cytokine measurement (**Section 5.3**). This cohort comprised 126 patients with ET from the PT-1 multicentre trial (see **Section 2.1**), who had undergone targeted genome sequencing as described in sections 3 and 4 and for whom long-term follow-up data was available. There thus was an intersection of 241 patients who both underwent targeted genome sequencing as described in sections 3 and 4, and who underwent quantification of serum cytokines using the 10-plex panel. This allowed us to model the effect of somatic mutations on the concentrations of the 10 measured cytokines, since certain mutations may be associated with differential cytokine secretion, or conversely, different mutant clones may have a selective advantage under certain micro-environmental conditions. After correction for diagnosis, age and sex, there were surprisingly few differences in cytokine profiles across patient genotypes. One exception was IP-10 (Figure 31), which was 1.5-fold higher in JAK2-mutated patients (p-value<0.001), 1.9-fold higher in JAK2-mutated patients with 9pUPD compared to those without (p-value=0.002), 2-fold higher when JAK2-mutations were found in combination with TET2-mutations (p-value=0.005) and correlated with JAK2 allele burden ($r^2=0.46$, $p<0.001$). A similar pattern of IP-10 secretion was also seen in JAK2 transgenic mice and those cross-bred with TET2-knock out mice (work done by Dr Juan Li, Miriam Belmonte and Dr David Kent).

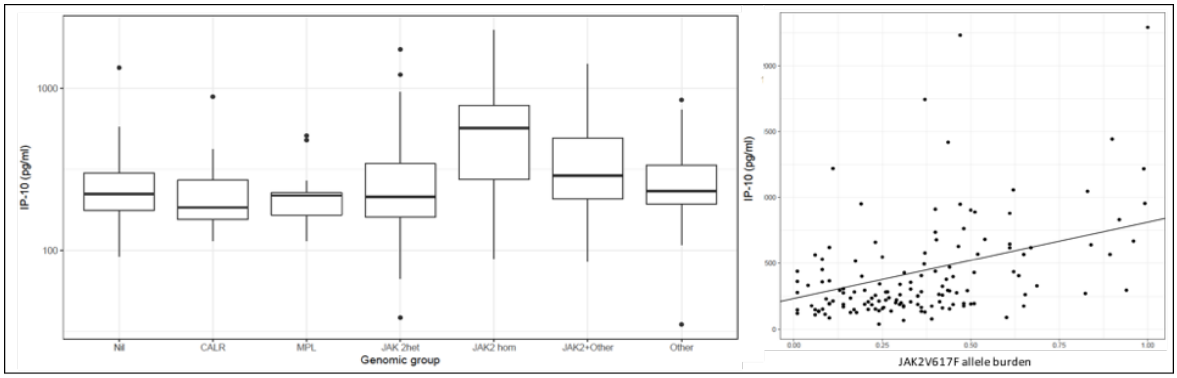


***Figure 31**: Correlates of IP-10 levels. (left) IP-10 concentrations across different genomic groups. JAK2 het=heterozygosity alone, JAK2hom=JAK2 homozygosity alone, JAK2 other includes JAK2 mutated patients with additional TET2 or DNMT3A, Other includes the TP53 and chromatin-spliceosome groups. (right) correlation of IP-10 levels with JAK2 burden.*

6.3 Prognostic correlates of pro-inflammatory cytokines in MPNs

6.3.1 Diagnostic cytokine levels

Previous studies have highlighted a prognostic role for a number of pro-inflammatory cytokines in MF. Outcome data was available for 182 of the original cohort and therefore this question could be evaluated both for MF and chronic phase patients. Unfortunately, this cohort had only 8 documented transformation events and 25 deaths. Cox proportional hazards modelling, incorporating clinical and demographic variables, suggested a prognostic role for GRO-alpha as a predictor for an increased risk of transformation from chronic phase to myelofibrosis or AML (p=0.002), although there were only a small number of events

The prognostic role for GRO-alpha was therefore evaluated in the 126-patient PT-1 cohort, in whom genomic variables could also be taken into account. This cohort had a median follow-up of 11.7 years (with a range of 2 months to 43 years from diagnosis) and was enriched for transformation events, with 26 AML transformations, 30 myelofibrotic transformations and 69 deaths. In this analysis samples were diluted 1:5 as it was noted that in some cases GRO-alpha levels exceeded the range covered by the standard curve.

Overall, the addition of cytokine quantification to clinical and genomic data improved the concordance of predictive models for overall survival and progression (Figure 32). GRO-alpha, as well as IL-8 and IP-10 in particular, were selected (using variable selection methods described previously) as predictive of progression (p-values 0.007, 0.007 and 0.01 respectively).

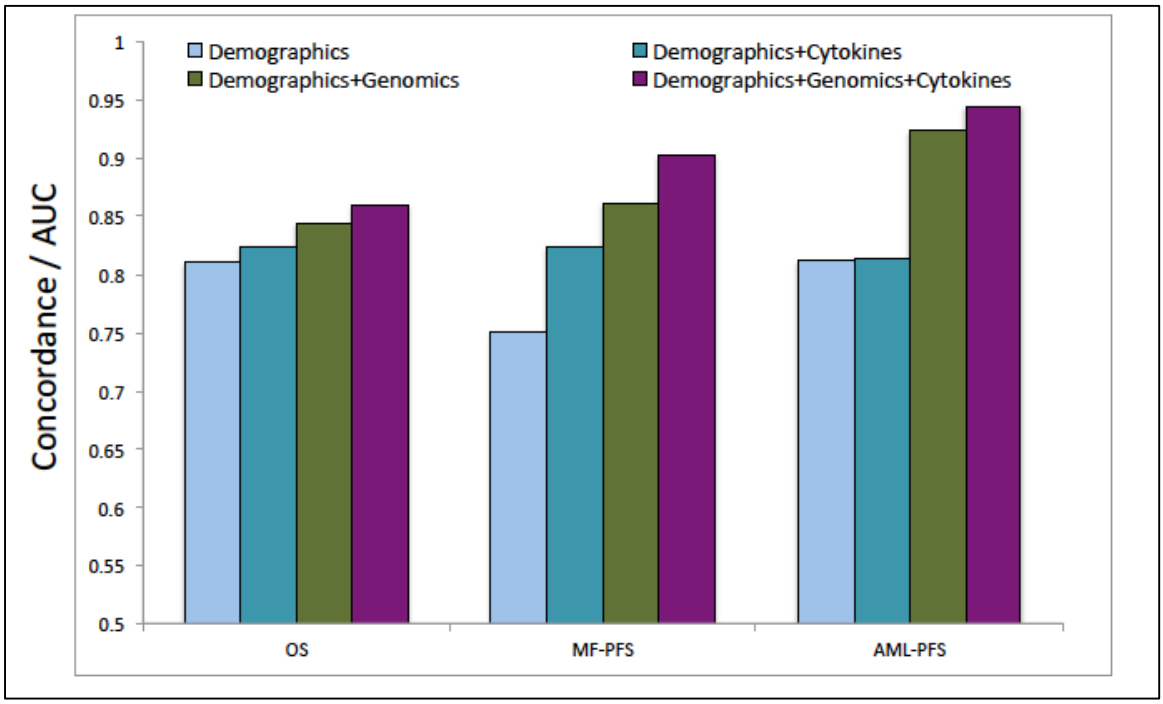


Figure 32: *Comparison of concordances for Cox proportional hazards modelling assess for the addition of cytokine and genomic data to a baseline model using age and sex alone, for overall survival, myelofibrotic transformation and AML transformation predictions.*

6.3.2 Longitudinal measurement of cytokine levels

As well as evaluating the predictive value of cytokine measurement at diagnosis, the value of longitudinal sampling was also assessed in 81 patients who had multiple samples and clinical follow-up. The rate of change in EGF strikingly correlated with transformation risk (p-value 0.008 in Cox proportional hazards modelling including age, sex and disease subtype), with the majority of transformations (~80%) observed in patients whose EGF levels decreased over time (Figure 33, stability defined as an absolute rate of change of <4pg/ml per year). Compared to patients with stable or increasing EGF levels, there was a 4.3-fold (95% confidence interval 1.7 - 10.9) increased likelihood of transformation to MF or AML if EGF levels dropped over the course of the longitudinal sampling, strongly indicating the utility of monitoring EGF levels during the course of disease.

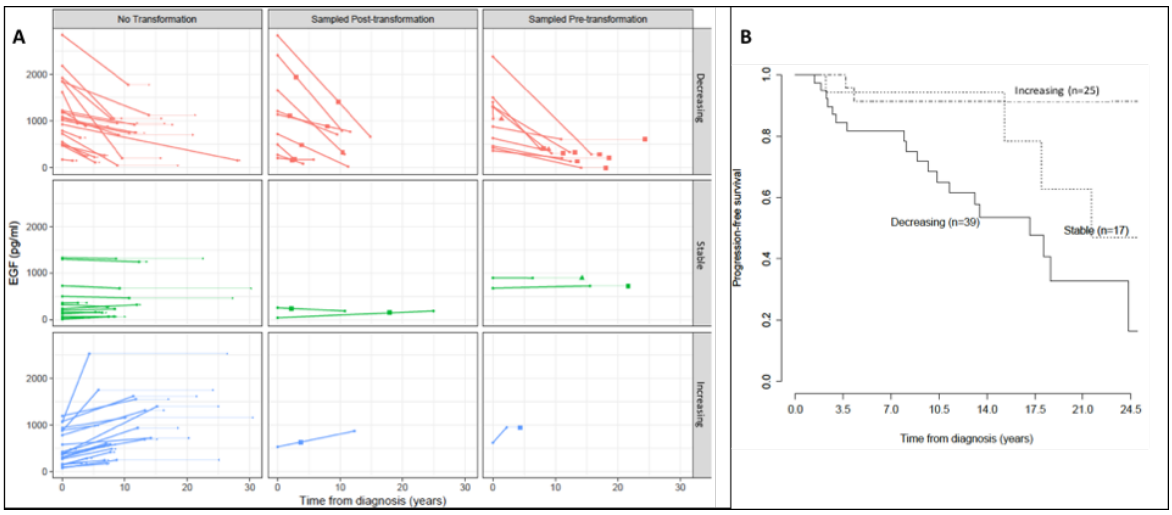


Figure 33: *Rate of change of EGF and outcome. (A) EGF levels over time. Change in concentration denoted by thick line joining the two sampling time-points, length of faded line indicates subsequent follow-up period until censoring or event. Square marker denotes time of MF transformation, Triangle denotes time of AML transformation. (B) Kaplan Meier analysis showing progression free survival, according to rate of change of EGF.*

6.4 Functional evaluation of MPN-associated cytokines

Although a number of the correlations between cytokine levels and clinical phenotype are quite convincing, the pathophysiology underlying them is not entirely clear. It is possible that the differences in cytokine levels are directly due to secretion by cells arising from the tumour clone (e.g. monocytes or macrophages) or they may be triggered indirectly by the effects of the tumour clone upon non-tumour cells in the bone marrow niche (whether stromal cells, or cells of the immune system). Conversely, particular micro-environmental conditions, existing independent of the tumour clone itself may promote expansion of the MPN clone, or influence the MPN phenotype, such that the same clone arising under two separate conditions would manifest as two separate phenotypes. It is also not clear to what extent the differences in cytokine profiles actually play a role in disease pathogenesis or simply are surrogate markers of the underlying myeloproliferation or, in other words are simply epiphenomena.

The 38-plex panel was used to measure the levels of the selected cytokines from supernatant from a megakaryocytic-differentiated induced pluripotent stem cell (IPSC)-derived cell line (supplied by Dr Thorsten Klampfl) demonstrate that these cells secrete EGF, IL-17, IL-1, IL-3, IL-8, IP-10, MIP-1a, MIP-1b, TNF-a, and GRO-alpha (at 10- to

100-fold higher levels than these other cytokines).

Work carried out by Dr Nina Friesgaard Øbro and Dr Miriam Belmonte using flow cytometric intra-cytoplasmic cytokine measurement indicates that stimulated CD14+ monocytes are the primary source of IP-10 in patients with ET and healthy controls, but that IP-10 production is additionally seen by populations of CD56+ monocytes, granulocytes and T cells in patients with PV and MF, with flow cytometric quantification of IP-10 correlating well serum levels measured with the 10- or 38-plex assays. T cells were found to be the main producers of IFN-γ and TNF-a, suggesting this was not direct secretion by the MPN clone, but potentially a reaction to it.

6.5 Discussion and Future work

The work presented here demonstrates that cytokine profiles differ significantly between patients with ET, PV and MF. Overall they form a continuum of increasing pro-inflammatory cytokines (including IP-10, TGF-alpha and TNF-alpha in particular), but patients with MF were found to have lower levels of GRO-alpha (which was highest in patients with ET), EGF and Eotaxin.

Further work is required to establish whether these correlations exist because of (a) cytokine secretion directly by the tumour clone, (b) secretion by non-tumour cells in response to the clone, (c) secretion by non-tumour cells independent but permitting or promoting tumour clone expansion, (d) secretion independent of the tumour clone but with other effects on disease pathogenesis (e.g. induction of fibrosis or angiogenesis) or (e) some combination of these. If secretion is by the tumour clone or in response to it, it also remains to be determined which cells or tissues are the targets and how this influences disease maintenance or progression, if at all.

The evidence presented here indirectly suggests IP-10 may be secreted by mutant monocytes, since IP-10 levels correlate closely with JAK2 clonal burden and the presence of homozygosity, while TNF-alpha and IFN-γ are more likely to be produced by T cells and therefore less likely to derive from the tumour clone itself. A number of cytokines, GRO-alpha in particular, were shown to be secreted by a megakaryocytic cell line, which is in keeping with the correlation between ET and high GRO-alpha levels. It is therefore likely that GRO-alpha levels are acting as a surrogate for megakaryocytic activity, although no clear correlation with platelet count was seen, and it is possible that this underlies the correlation seen between high GRO-alpha levels and risk of myelofibrotic

transformation. As GRO-alpha acts as a neutrophil chemo-attractant and has a role in angiogenesis, it is possible that it may stimulate fibrosis via other mechanisms.

Previous studies have suggested that the secretion IL-6, TNF-alpha and TGF-beta may provide a selective advantage to the mutant clone, while inhibiting normal haematopoiesis. Work done in the Kent lab suggests a similar role for IFN-γ (unpublished data), with increased colony growth seen with IFN-γ treated colonies from patients with MF, but inhibited growth in samples from patients with ET. Surprisingly, other than IP-10, none of the cytokines measured correlated significantly with tumour burden or the underlying mutational profile, which would appear to argue against secretion by tumour-derived cells themselves and potentially for a model in which cytokine secretion serves to permit or encourage expansion of the clone relative to wild type cells.

The mechanisms by which these cytokines may contribute to disease are difficult to determine and further in vitro and in vivo work is required to elucidate this. It is, for example, difficult to hypothesise why MF is associated with increased TGF-alpha levels but lower EGF levels (which here are shown to predict for MF progression) since both molecules have similar functions and act on the same receptor (epidermal growth factor receptor, EGFR). Further work might include comparing secretion profiles of tumour-derived megakaryocytes and monocytes to wild-type controls, assessing the effect of these cytokines on colony growth and differentiation, and assessing their effect on fibroblast function, including collagen production. Finally, the contribution of these cytokines to disease establishment and progression could be assessed in vivo through the use of cytokine-knockout mouse models along with transplant experiments using established JAK2 or CALR transgenic models of MPN.

Therefore, although much work remains to be done in order to understand the role of the micro-environment in MPN pathogenesis, the work presented here should hopefully form a strong basis for this through the identification of a number of candidate cytokines, and additionally identifies a number of candidate biomarkers for disease progression which could be validated in future studies.

7 Discussion

7.1 Introduction

The Philadelphia-negative myeloproliferative neoplasms offer a promising disease model for understanding how haematological malignancies originate and develop. They are relatively genomically simple, with few mutations per patient, and the majority of patients harbour JAK2, CALR or MPL mutations. Genomically, however, they overlap not only with each other but also with MDS and AML, and also with otherwise healthy patients who can be found to harbour expanded clones carrying JAK2 or other mutations (CHIP/ARCH). Clinically, patients can transition between these disease states: CHIP to an MPN, ET or PV to MF, MPN to MDS and MPN or MDS to AML.

It is therefore remains a challenge to determine which variables, genomic or otherwise, determine which clinical phenotype a patient may have (if any at all) and their risk of progression into another disease state.

7.2 The Genomic Landscape of MPNs

7.2.1 Overview of somatic genomic variation in MPNs

The results of this large sequencing study are consistent with previous sequencing studies in MPNs, with recurrent mutations seen across genes implicated in cell signalling, epigenetic and transcriptional regulation and mRNA splicing. As well as JAK2, CALR, MPL, CBL and SH2B3, which are well understood to be involved in cellular responses to Epo, Tpo and G-CSF with down-stream JAK-STAT signalling, mutations in GNAS and GNB1 were also seen, the latter restricted to the K57 hot-spot. Mutations in G proteins are only recently described in a small number of cases of myeloid and lymphoid malignancies as well as in cases with clonal haematopoiesis[249,250], and have been associated with resistance to oncogenic kinase inhibition. Although only a small number of cases were seen in this cohort, there did appear to be an association with poorer outcomes, and therefore this may represent a clinically important pathway, disrupted in a subset of patients.

Mutations in TET2, DNMT3A, ASXL1, EZH2 were frequently found, as were mutations in splicing components, SRSF2, ZRSR2, U2AF1 and SF3B1 in a smaller number. SF3B1, however, appears to have a different functional role in MPNs than the other mutated

spliceosome genes which tend to occur later, are associated with increasing age and disease burdens, as well as with significant cytopenias and poor risk disease across both chronic phase and myelofibrotic disease. This is not the cases for SF3B1, which generally occurs early, and is associated with anaemia alone, but not with as significant a risk of disease transformation. This may reflect the predominant effects of SF3B1 on iron homeostasis and erythropoiesis, in contrast to the more significant disruption of stem cell function seen in mouse models with the other spliceosomal components.

Mutations of TP53 were recurrent, as previously reported, but PPM1D, a negative regulator of TP53, was also shown to be mutated, with mutations showing a similar pattern to those seen in other cancers[221,251]. Although previous reports have hypothesised that these mutations represent germ-line mosaicism, evidence is presented here supporting their somatic nature. They are demonstrated here to be late events, often occuring secondary to JAK2 or CALR mutations, but also seen in a small number of cases as being isolated events. Although occurring together with TP53 or 17p UPD/deletions in a small number of cases, no statistically significant enrichment was seen, and these mutations were not significant predictors for AML transformation, unlike those in TP53 (or 5q/17p changes). These mutations appear to be enriched in patients with chemotherapy exposure, as demonstrated both here in relation to hydroxycarbamide exposure and other studies, and may be associated with resistance to therapy[252]. However, their presence at diagnosis or early in disease, prior to any cytotoxic exposure, is seen in a subset of patients here. It may be that screening for these mutations may be helpful in predicting treatment responses, something that could be examined in future clinical trials.

Frameshift/nonsense MLL3 mutations, previously reported only in a very small number of AML cases[224], are here demonstrated to be recurrent in MPNs, with no clear phenotypic or prognostic associations. These mutations however are enriched in triple-negative patients, and can occur in isolation, suggesting they may be sufficient to cause clonal expansion and disease, a hypothesis that needs to be tested in vitro, as well as screened for in a wider, triple-negative population. Although a number of other putative candidates were suggested by these data, there were no clear novel phenotypic drivers discovered in this cohort.

Recurrent chromosomal changes were also seen across 14 chromosomes, most commonly on 9p, associated with JAK2 mutations, leading to their homozygosity. UPD or copy-number changes on chromosomes 1, 4, 7, 11, 12, 17 and 19 were also seen recurrently and

strongly associated with mutations in MPL (including one non-canonical variant), TET2, EZH2 (although also seen independently), CBL, SH2B3, TP53, and CALR, leading to loss of heterozygosity for these mutations. UPD was also seen affecting chromosomes 13, 14 and 18, with no associated mutations detected, and no recurrent SNPs (although relatively few were genotyped in this study). Further work is required to determine whether UPD on these chromosomes is associated with unrecognised somatic mutations and/or with loss of heterozygosity for functionally relevant germ-line SNPs, as well as the functional consequences of 5q and 20q deletions and trisomy 8.

7.2.2 Sufficiency of phenotypic driver mutations for MPN development

Previous studies of mouse models of MPNs have suggested that while JAK2 and CALR mutations are sufficient to give rise to a myeloproliferative phenotype, they do not necessarily provide a stem cell advantage, leading in some cases to a failure to give rise to sustainable disease in competitive transplant experiments. Furthermore, JAK2 mutations are found in a proportion of the general public in the absence of any overt haematological abnormality. In the proportion that do progress to develop an MPN, there is often a lag-time of many years. One hypothesis is that mutations in other genes are required to provide a sufficient competitive advantage and to initiate disease. It is notable that many of the additional mutations seen in this and other studies, including TET2, DNMT3A, ASXL1 and EZH2, are associated with expansion of the stem cell pool, and therefore are likely to offer a relative clonal advantage to the mutant clone.

The data presented here do not support this hypothesis, since ~50% of JAK2, CALR or MPL-mutated cases were not found to have mutations in other genes, or to have chromosomal abnormalities. However, this study is limited in that only a targeted set of genes were sequenced, and only their coding regions. and it is therefore possible that mutations crucial for MPN development are present, but not captured by this study.

A number of pro-inflammatory cytokines were found in this and other studies, to be raised in MPNs, particularly in cases of MF. These include TNF-alpha and IL-6 which have been found to promote the expansion of JAK2-mutant clones while inhibiting normal haematopoiesis, and unpublished data suggests interferon-gamma may also preferentially support the growth of mutant clones. These cytokines may in some cases be produced or induced by the tumour clone, but are also associated with ageing, and also with a large number of inflammatory conditions. It is therefore feasible that the probability of a JAK2 clone expanding and giving rise to disease is influenced by changes in the bone marrow

micro-environment induced by the clone itself, by the ageing process or by other inflammatory processes, as well as differences in the relative fitness of the wild-type clones. These differences may relate to the genetic background of the patient, including JAK2 haplotype in particular, to other disease processes, or toxin exposure, e.g. smoking (rates of which are higher in patients with PV than unaffected controls) or chemotherapy exposure. It is possible that the threshold required for an increase in erythropoiesis/thrombopoiesis to reach clinically detectable levels will depend on other factors (concurrent iron deficiency, smoking or alcohol use, or germ-line variation in baseline haemoglobin or platelet counts) that vary between patients.

Given the heterogeneity of mutations and cytokine profiles observed, rates of progression to MPN and age of presentation, it is reasonable therefore to hypothesise that multiple paths exist leading from JAK2-mutated CHIP to MPN, such that additional mutations or stimulation from pro-inflammatory signals are needed to overcome reduced fitness in some contexts but not in others. While attractive, this model is not entirely borne out by the data here. For example, MPL mutations, which have not been reported in the general population and therefore are likely to carry a high probability of resulting in an MPN phenotype, tend to be associated with a higher mutation burden than JAK2 and CALR mutations. Furthermore, increasing mutation burden and pro-inflammatory states, rather appearing to exist as alternative pathways to MPN tend to co-occur, each correlating with increasing age and with myelofibrosis.

This highlights the need for additional studies addressing in detail the differences in germ-line variants, somatic mutation rates, inflammatory signals and chemical/radiation exposure between patients with MPNs and those with CHIP/ARCH.

7.3 Determinants of phenotype in JAK2-mutated chronic phase patients

The comprehensive clinical and genomic description of this cohorts allows us to examine the factors influencing the development of PV rather than ET in patients with JAK2 V617F mutations.

7.3.1 Demographic, Micro-environmental and Genomic factors

Male sex and increasing age were associated with PV phenotype. The association with male sex may reflect intrinsic germ-line differences affecting stem cell function, but may also be a function of differences in hormones, iron stores, inflammatory signals or environmental factors (e.g. toxin exposure). The correlation with increasing age supports

the hypothesis that age-related changes in the bone marrow may lead to selection of different stem/progenitor populations, and may be supporting the expansion of JAK2 clones, in keeping with multiple lines of evidence showing an association between JAK2 burden/dosage and erythrocytosis. Indeed, a number of pro-inflammatory cytokines, several of which are implicated in supporting JAK2-mutant clone growth, were higher in patients with PV than those with ET, including TNF-alpha, TGF-alpha, IP-10, and to a lesser extent IL-6 and IL-8. Differences in IP-10 secretion are particularly of interest, since they correlate closely with JAK2 burden, and also because of previous work implicating differences in interferon signalling seen between patients with ET and PV[45]. Conversely, much higher GRO-alpha levels are associated with ET. However, further work is required to investigate causality here, since high JAK2 burden disease may be responsible for higher pro-inflammatory cytokine levels rather than the other way around, and GRO-alpha levels may be a surrogate for increased megakaryocytic activity rather than themselves causing thrombocytosis (or lower haemoglobin concentrations).

The type of JAK2 mutation also appears to correlate with phenotype, most significantly with exon 12 mutations, however a number of non-canonical JAK2 mutations also appear to be restricted to particular phenotypes. This may reflect differences in EPOR or MPL binding or the strength of signalling, with different degrees of JAK2, STAT or ERK phosphorylation.

NFE2 mutations were also enriched in patients with PV, as were STAG2 mutations in a small number, and tended to co-occur with 9pUPD. As noted previously, this is in keeping with previous studies showing an association between NFE2 mutations and expanded erythropoiesis[114]. The expression of NFE2 is regulated by pro-inflammatory signals, including IL-1beta, and in turn its expression (increased either due to these signals, or mutation) is associated with increased IL-8 production, potentially causing a feedback loop, sustaining disease[253]. Another genomic variable strongly associated with MPN phenotype in JAK2-mutated cases was rs11042125, which has been associated with haemoglobin concentrations in the general population[205], suggesting germ-line determinants of haematological parameters also further modify the phenotype in patients with MPNs.

7.3.2 JAK2 mutation order and V617F homozygosity

The data presented here confirm a strong correlation between JAK2 clone size and JAK2

homozygosity with PV phenotype, with detectable 9pUPD a rare event in ET patients. They also confirm the association between the order of JAK2 mutation acquisition and ET/PV phenotype in the ~50% of patients who carry mutations in addition to those of JAK2. Earlier acquisition of JAK2 mutations ("JAK2 first"), and therefore their presence in the dominant clone, was shown to associate with a greater incidence of 9pUPD and with a PV phenotype. This relationship was the case even in cases lacking detectable 9pUPD, suggesting this was not the sole determinant, and appeared to hold regardless of partner mutation (TET2, DNMT3A or other mutations). The possible mechanisms underlying this have been discussed in earlier sections, but it is again tempting to hypothesise a role for micro-environmental changes in determining these phenotypic differences, with an earlier and large JAK2 clone leading to a longer period of bone marrow modelling and cytokine secretion supporting out-growth of the JAK2 clone

7.3.3 JAK2 haplotype and V617F homozygosity

A significant association between the JAK2 46/1 haplotype and the presence of 9pUPD and PV phenotype was also seen in this study. This is an intriguing finding which raises a number of questions about MPN pathogenesis. It is consistent with previous reports from other groups that reported associations between PV and SNPs in linkage dysequilibrium with 46/1 haplotype SNPs[140,147]and showing more rapid increases in JAK2 allele burden over time in those carrying the 46/1 haplotype[144,145,160]. There are number of possible explanations for these associations which are not necessarily mutually exclusive. The presence of the 46/1 haplotype may be associated with greater regional instability, as has been previously suggested[230], leading to a greater probability of both JAK2 mutation and mitotic recombination leading to 9pUPD (or increasing the possibility of bi-allelic JAK2 mutations); an additional competitive advantage may be imparted to the JAK2-mutant clone, when the mutation is acquired on the 46/1 background (although changes in JAK2 expression or splicing have not been demonstrated in previous studies[141]), and outgrowth of the clone increases the probability of 9pUPD occurring; and acquired homozygosity for the 46/1 haplotype as a result of 9pUPD may itself offer an additional clonal advantage.

A clonal advantage for the JAK2 46/1 haplotype in its own right is supported here by the demonstration that the haplotype is not just enriched in cases with JAK2 mutations, but also in cases with MPL, and to a lesser extent, CALR mutations. Previous studies have also demonstrated a small number of cases where acquired homozygosity for 9pUPD occurred prior to the acquisition of the JAK2 mutation[7,254], suggesting that acquired

homozygosity for germ-line variants (with 46/1 haplotype as one reasonable candidate) may offer an advantage to the clone.

Furthermore, as well as its association with MPN-related mutations, the JAK2 46/1 haplotype has been shown to be enriched in patients with Crohn's disease, and associated with increased rates of DNA damage in these patients[255–257]. It has also been shown to be associated with normal karyotype acute myelomonocytic leukaemia and with adverse outcomes overall in normal karyotype AML[258]overall suggesting that the haplotype may be associated with abnormal myelomonocytic activation and aberrant immune responses[259], and presenting another pathway for germ-line – environmental interaction.

7.3.4 9pUPD and other genes on 9p

Additionally, a number of SNPs implicated in variation in haematological parameters including rs409801 and rs385893, associated with platelet counts, are also located on 9p. It is possible that homozygosity for 9pUPD, leading to acquired homozygosity for these variants, in turn contributing to further skewing of blood counts and alteration of the resultant MPN phenotype. NF-IB is also located on 9p, centromeric to JAK2 (Figure 16 (C)). Earlier reports of 9pUPD demonstrated markedly increased expression of NF-IB in patients with 9pUPD and hypothesised that this offers a protective advantage to the clone from the inhibitory effects of TGF-beta[7].

Intriguingly, a cluster of interferon genes is also present on 9p21.3 (Figure 16 (D)). This is a locus that has been implicated in cardiovascular risk (coronary artery and aortic disease) across a large number of studies and in particular, rs4977574, an intergenic SNP associated with CDKN2A and CDKN2B has been identified as a significant predictor of risk[260,261]. Unfortunately, SNPs in this region were not covered by the bait sets used in this study, but this would be worth following up in future studies, particularly because of the variation seen in thrombotic risk in MPN patients and the need for risk stratification, as well as the differential interferon responses noted in patients with ET or PV[45]. Preliminary data presented here also suggest that the length of chromosome 9p involved in 9pUPD varies between patients with ET and PV. It is possible that differences in involved genes and functional SNPs affected by 9pUPD underlie some of the differences in interferon response and phenotype.

Finally, PD-L1 (CD274, Figure 16 (B)) lies adjacent to JAK2. It is of particular interest in haematological malignancies as its expression has been shown to be increased in lymphoma in conjunction with gains/amplification of 9p, offering a mechanism for immune evasion. It therefore is another candidate gene affected by 9pUPD, and indeed recent data show increased PD-L1 expression in JAK2-mutants (although the mechanism underlying this is unknown), associated with immune evasion in cell line and murine models.

In summary therefore, the phenotypic differences seen among patients with JAK2-mutated chronic phase disease appear to depend on a number of factors including (a) age and sex, (b) JAK2 allele burden, (c) the presence of 9pUPD, (d) 46/1 haplotype +/- acquired homozygosity due to 9pUPD (e) other genomic variation on 9pUPD +/- acquired homozygosity due to 9pUPD, (f) genomic variation at other sites (e.g. rs11042125), (g) additional somatic mutations (potentially in NFE2 and STAG2) and (h) levels of pro-inflammatory cytokines. Further comprehesive analyses examining candidate SNPs along 9p and elsewhere in conjunction with the presence of 9pUPD and other somatic changes are required to fully elucidate these interactions.

7.4 Pathogenesis of Myelofibrosis and Blast Transformation

7.4.1 Myelofibrosis and recurrent somatic mutations

A recurring feature of the analyses presented in this thesis is the cluster of genes including ASXL1, EZH2, SRSF2 and U2AF1. ASXL1 mutation is often seen in EZH2, SRSF2 or U2AF1-mutated cases, but these mutations are not seen together. This family of mutations was identified by clustering using a Dirichlet-process based Bayesian clustering algorithm, and was also identified using this method when the data was combined with an independent dataset from patients with MDS, and also is associated with mutations in ZRSR2, IDH1/2, NRAS, CBL and GNAS, as well as 7q deletions/UPD.

This cluster is strongly enriched in patients with MF, as well as in patients with chronic phase disease subsequently transforming to MF. It is possible that mutations in these genes present alternative pathways for the development of MF, but the mutual exclusivity of EZH2, U2AF1 and SRSF2 both in this cohort and others points to functional redundancy and the existence of common mechanisms, as do the similarity in mutation partners for these mutations (ASXL1 in particular).

The mechanisms that may underlie this have been discussed in previous sections, but

common to these mutations appear to be the following features: an association with anaemia +/- thrombocytopenia in keeping with their association with dysplasia and ineffective haematopoiesis; expansion of the stem cell pool in murine models, often with maturation arrest; dysregulation of H3K27 methylation; dysregulation of genes regulating stem cell quiescence and differentiation, including genes of the HOXA cluster; the development of bone marrow fibrosis; poor risk disease in both chronic phase and myelofibrosis, with an increased risk of AML transformation.

Unfortunately, no correlation was seen between mutations in this set of genes and cytokine secretion, after correction for MPN subtype, going against the hypotheses that either these mutations lead to a pro-inflammatory/pro-fibrotic environment or that such an environment is supportive for the expansion of these mutant clones. However, these hypotheses cannot be discounted altogether as these same mechanisms may be present in other cases of MF in the absence of these mutations, obscuring a possible correlation.

The effects of this cluster of mutations appear to transcend the background on which they occur (whether in JAK2-, CALR- or MPL-mutated or triple negative disease) as well as the MPN subtype they are found in, justifying the proposal that they exist as an entitity in their own right. As noted above, this group of mutations is also frequently seen in MDS and AML, where it is also associated with poorer outcomes[189,214].

7.4.2 Determinants of blast transformation

Overlap was seen between the predictors of AML transformation from CP and MF. Ultimately these two cohorts were combined for the prediction of AML transformation in the final multi-stage model.

ASXL1, EZH2, U2AF1 and SRSF2 are also strongly correlated with risk of AML transformation. Other predictors include TET2 and DNMT3A, which are also associated with expansion of the stem cell pool in mouse models. Although the associated risk carried by these mutations is small, it is likely to be clinically relevant as they occur at a greater frequency than other mutations. GNAS, NRAS, RUNX1 and trisomy 9 are also associated with an increased risk of transformation, although the mechanisms underlying this are less clear.

TP53 mutations carry the strongest risk of AML transformation, and are commonly found with 5q and 17p changes, themselves associated with increased AML risk. This set of changes also was found to define a separate cluster in Bayesian clustering analysis. Again,

the phenotypic consequences of this group (with a >10-fold increased risk of AML transformation) transcend both the background on which these mutations occur, and the MPN subtype the patient has been classified into. Furthermore, the size of the TP53 clone at the time of sampling (near to diagnosis in the majority of cases) does not appear to correlate with the risk of transformation, suggesting its presence alone is sufficient to mark the patient as being at high risk and suggesting either that these clones expand over time, regardless of their size at MPN diagnosis, or that AML clones may arise even from small TP53 clones, gaining an advantage over the dominant clone.

7.5 Ontological status of MPN entities

There has been much debate regarding the diagnostic criteria for ET, PV and primary MF. Controversies include the correct thresholds for determining a raised red cell mass, whether haemoglobin or haematocrit should be used, whether ET and PV are distinct biological entities or sit on a continuum, and whether a continuum exists between ET and MF, or between MF and a pre-fibrotic state (pre-fibrotic MF). These issues are complicated by confounding variables (such as plasma volume or gender) and inter-observer variability when defining histological grading or subgroups[192,262–264].

The analyses reported above present a complex picture of the myeloproliferative neoplasms, with numerous lines of evidence suggesting that the traditional division into ET, PV and MF is an over-simplification that does not necessarily respect the underlying biological features of these conditions.

The Bayesian clustering algorithm defines a number of genomically defined entities, based on genomic data alone. The examples given above, of the spliceosome-chromatin and TP53-17p-5q groups demonstrate that these entities transcend clinical MPN entities but also are strongly correlated with the risk of disease transformation and with overall survival, with a classification based solely on these entities showing better predictive ability (assessed using concordance) than one using the MPN subtype alone. Furthermore, in multivariate, multi-stage modelling, the removal of the distinction between ET or PV did not significantly worsen the performance of the model, once genomic, demographic and clinical data are taken into account.

Furthermore, the since the spliceosome-chromatin group is enriched both in patients with MF and those that subsequently transform to MF and in both cases with poor outcomes, it appears arbitrary to classify these patients into two separate diagnostic entities based on the disease state in which they present, and furthermore to classify the subgroup of ET

patients with spliceosome-chromatin group mutations (and high risk of transformation and poor OS) into the same classification as ET patients with an isolated JAK2 mutation or no detectable mutations, for example. The division of remaining groups, defined predominantly by the presence of JAK2, JAK2 homozygosity, CALR or MPL, identifies patients with more benign disease, and as previously shown, correlates with differences in haematological variables and thrombotic risk[42,265].

A genomic classification, such as this one, offers the advantages of objectivity and reproducibility, and, for the reasons described above, may improve the ability to make predictions about possible outcome based on diagnostic classification alone. Furthermore, given it is based on the genomic lesions that are present, may aid the targeting of therapeutic agents by identifying specific targets or identifying specific patient groups that may or not benefit from specific treatments. This classification also identifies entities that may transcend the boundaries between MPN, MDS, MPN/MDS overlap syndromes and AML, and it may be more appropriate for a patient's management to be targeted according to genomic boundaries rather than specifically on criteria such as thrombocytosis, dysplasia or bone marrow fibrosis that may results in patients with a similar biological process (e.g. SRSF2 and ASXL1 mutated disease) being classified into separate diagnostic categories and being treated in different ways as a result. Conversely, this genomic classification also picks out a class of patients with no detectable mutations, who carry a low risk of adverse outcomes, and therefore again should not necessarily be grouped together, and treated the same as, a cohort of patients with confirmed clonal disease.

However, it is evident from the analyses presented above that both these genomic entities and existing clinical entities do not fully account for the phenotypic heterogeneity seen in patients with MPNs. As argued here, and in a previous review[262], it may be more accurate to view individual patients as sitting within a continuum (or phase space) with their phenotype depending on many variables such as age, sex, somatic mutations, the ordering and clonal composition of these mutations, chromosomal changes, germ-line SNPs (and the acquisition of homozygosity for these SNPs in some cases), levels of pro-inflammatory cytokines, iron stores, among others.

These variables then determine factors fundamental to the disease phenotype and behaviour, such as the degree of erythrocytosis (with early JAK2 mutations, NFE2 or JAK2 exon 12 mutations, 46/1 haplotype, 9p UPD at one end, late JAK2, CALR or MPL

mutations, rs11042125 genotype, spliceosomal or chromatin mutations at the other); the degree of stem cell expansion or dysplasia (including age, inflammatory states, DNMT3A, TET2, TP53, spliceosomal or chromatin mutations) and the degree of fibrosis (including age, inflammatory states, spliceosomal or chromatin mutations). These variables also determine a patients of moving within this phase space, for example towards a more myelofibrotic phenotype, or to AML.

This model therefore argues for both a clinical and genomic assessment, leading to an individualised approach to both disease classification and risk assessment. However, a wider study that includes patients with overlap syndromes, MDS, AML and CHIP/ARCH is required in order to fully understand the biological differences that underlie these phenotypic entities and determine a patients transition between states.

7.6 Prognostic utility of genomic data

A number of predictive scores for myelofibrosis are validated in this study, including HMR, IPSS, DIPSS and the use of CALR/ASXL1 alone, and here we show that they can, in some cases, be extrapolated to chronic phase patients. Presently available prognostic scoring systems have a number of advantages including their ease of use and the requirement for only a few known variables, however they have a number of limitations and flaws, which include:

- The lack of existing prognostic modelling for patients with chronic phase disease. Prognostication in these patients carries an additional dimension as they can transform to myelofibrosis.

- These models work by assigning scores of 0-2 for each variable, either for presence/absence or by dichotomising a continuous variable. This results in loss of available information - e.g. for a cut-off of 65 years, assuming all else is equal, a patient of 20 is assumed to have the same risk as one of 64, and one of 66 as one of 90, while the 64 and 66 year olds are treated as having different risks. Furthermore the standardisation of each variable to single integer value of 1 or 2 also may not accurately reflect the risk associated with each variable. Potential information is lost further when patients with different totals are grouped together.

- Previous studies generally chose a small number of variables, generally selected from a relatively small initial pool of variables. While using a score with only few

variables improves simplicity of the model, variable selection again leads to loss of information and then to overestimation of the effect sizes of the selected variables when models are re-built using only the selected variables.

- Cytogenetic analyses are used to detect chromosomal amplifications/losses, but copy-number neutral events are not included in any analyses.

- These analysis generally classify patients into 3 or 4 risk groups. Predictions are not specific to a given individual, and therefore comparisons between patients within a risk group cannot be further refined, and predictions for a particular disease outcome (e.g. AML transformation vs. non-AML death) or particular time-point cannot be made (or need to be inferred from the behaviour of the risk group as a whole). In other words, while modelling is used to select variables, a predictive model in its own right is not created, but rather patients are assumed to have a similar risk to that of those in the relevant prognostic category.

On this basis, we have presented here an integrated prognostic model that encorporates data from patients with both CP and MF, using demographic, clinical and genomic data, accounting for transitions between disease states and that enables predictions to be provided for specific outcomes for specific time-points. These predictions are therefore amenable to direct comparison against actual outcomes. An additional benefit of this approach is that risk groups are not defined a priori, but rather continuous predictions are given, which allows the model to be tailored as required if categorisation into risk groups is required. In Appendix 5 we provide estimates of the number of patients needed to test using this model to define patients groups meeting particular thresholds. Analyses such as this may then inform clinical trial design (e.g. identifying a target high-risk cohort for a trial of a novel treatment) or day-to-day patient management (e.g. identifying a very low risk cohort that do not require regular clinic follow-up or intervention).

We have shown that these models perform well when compared to those lacking full genomic characterisation and to models designed using variable selection methods. They also perform at least as well at risk stratifying patients. Even if not adopted, they also inform the selection of variables for more simplified scoring systems. For example, it is clear that the inclusion of a number of more rare variables, such as TP53 or NRAS, may result in significantly different predictions being made in individual cases.

As discussed in an earlier section, there are a number of ways in which the modelling

presented here could be improved, including a larger knowledge bank, inclusion of additional variables, and alterations in model fitting. Given the data presented regarding GRO-alpha and EGF trends, it may be that cytokine profiling can offer an additional and dynamic means of prognostication that could be built into existing models.

7.7 Conclusion

The majority of myeloproliferative neoplasms display a narrow repertoire of mutations in genes involved in erythropoietin or thrombopoietin signalling with a long tail of less frequent mutations. Despite this, we have shown here that a better understanding of the additional somatic events (mutations or chromosomal changes) and germ-line genomic background can inform our understanding of the processes giving rise to the myeloproliferative phenotypes in these patients, and can help us to accurately predict the risk of disease transformation. Striking differences in cytokine profiles are also seen They also argue for a personalised medical approach, where these variables, together with clinical and demographic variables, allow for better characterisation of the disease phenotype and associated risk, rather than assigning an individual to a broad diagnostic group or risk category.

Further work examining the role of somatic and germ-line variation, the interaction between the two, and the contribution of micro-environmental factors will be required in order to understand the phenotypic heterogeneity of these conditions and their relationship to other related clonal haematopoietic disorders.

References

1. Barbui T, Thiele J, Gisslinger H, et al. The 2016 WHO classification and diagnostic criteria for myeloproliferative neoplasms: document summary and in-depth discussion. Blood Cancer J 2018;8(2):15.

2. Adamson JW, Fialkow PJ, Murphy S, Prchal JF, Steinmann L. Polycythemia Vera: Stem-Cell and Probable Clonal Origin of the Disease. N Engl J Med 1976;295(17):913–916.

3. Fialkow PJ, Faguet GB, Jacobson RJ, Vaidya K, Murphy S. Evidence that essential thrombocythemia is a clonal disorder with origin in a multipotent stem cell. Blood 1981;58(5):916–919.

4. Testa JR, Kinnealey A, Rowley JD, Golde DW, Potter D. Deletion of the long arm of chromosome 20 [del(20)(q11)] in myeloid disorders. Blood 1978;52(5):868–877.

5. Nacheva E, Holloway T, Carter N, Grace C, White N, Green AR. Characterization of 20q deletions in patients with myeloproliferative disorders or myelodysplastic syndromes. Cancer Genet Cytogenet 1995;80(2):87–94.

6. Rege-Cambrin G, Mecucci C, Tricot G, et al. A chromosomal profile of polycythemia vera. Cancer Genet Cytogenet 1987;25(2):233–245.

7. Kralovics R, Guan Y, Prchal JT. Acquired uniparental disomy of chromosome 9p is a frequent stem cell defect in polycythemia vera. Exp Hematol 2002;30(3):229–236.

8. Dameshek W. Editorial: Some Speculations on the Myeloproliferative Syndromes. Blood 1951;6(4):372–375.

9. Baxter EJ, Scott LM, Campbell PJ, et al. Acquired mutation of the tyrosine kinase JAK2 in human myeloproliferative disorders. The Lancet 2005;365(9464):1054–1061.

10. James C, Ugo V, Le Couédic J-P, et al. A unique clonal JAK2 mutation leading to constitutive signalling causes polycythaemia vera. Nature 2005;434(7037):1144–1148.

11. Kralovics R, Passamonti F, Buser AS, et al. A Gain-of-Function Mutation of JAK2 in Myeloproliferative Disorders. N Engl J Med 2005;352(17):1779–1790.

12. Levine RL, Wadleigh M, Cools J, et al. Activating mutation in the tyrosine kinase JAK2 in polycythemia vera, essential thrombocythemia, and myeloid metaplasia with myelofibrosis. Cancer Cell 2005;7(4):387–397.

13. Silvennoinen O, Hubbard SR. Molecular insights into regulation of JAK2 in myeloproliferative neoplasms. Blood 2015;125(22):3388–3392.

14. Dusa A, Mouton C, Pecquet C, Herman M, Constantinescu SN. JAK2 V617F Constitutive Activation Requires JH2 Residue F595: A Pseudokinase Domain Target for Specific Inhibitors. PLoS ONE 2010;5(6):e11157.

15. Sanz Sanz A, Niranjan Y, Hammarén H, et al. The JH2 domain and SH2-JH2 linker regulate JAK2 activity: A detailed kinetic analysis of wild type and V617F mutant kinase domains. Biochim Biophys Acta BBA - Proteins Proteomics 2014;1844(10):1835–1841.

16. Lu X, Levine R, Tong W, et al. Expression of a homodimeric type I cytokine receptor is required for JAK2V617F-mediated transformation. Proc Natl Acad Sci U S A 2005;102(52):18962–18967.

17. Wernig G, Gonneville JR, Crowley BJ, et al. The Jak2V617F oncogene associated with myeloproliferative diseases requires a functional FERM domain for transformation and for expression of the Myc and Pim proto-oncogenes. Blood 2008;111(7):3751–3759.

18. Hookham MB, Elliott J, Suessmuth Y, et al. The myeloproliferative disorder–associated JAK2 V617F mutant escapes negative regulation by suppressor of cytokine signaling 3. Blood 2007;109(11):4924–4929.

19. Yan D, Hutchison RE, Mohi G. Critical requirement for Stat5 in a mouse model of polycythemia vera. Blood 2012;119(15):3539–3549.

20. Mnjoyan Z, Yoon D, Li J, Delhommeau F, Afshar-Kharghan V. The effect of the JAK2 V617F mutation on PRV-1 expression. Haematologica 2006;91(3):411–412.

21. Dillon M, Minear J, Johnson J, Lannutti BJ. Expression of the GPI-anchored receptor Prv-1 enhances thrombopoietin and IL-3-induced proliferation in hematopoietic cell lines. Leuk Res 2008;32(5):811–819.

22. Kralovics R, Teo S-S, Buser AS, et al. Altered gene expression in myeloproliferative disorders correlates with activation of signaling by the V617F mutation of Jak2. Blood 2005;106(10):3374–3376.

23. Mutschler M, Magin AS, Buerge M, et al. NF-E2 overexpression delays erythroid maturation and increases erythrocyte production. Br J Haematol 2009;146(2):203–217.

24. Bogeska R, Pahl HL. Elevated Nuclear Factor Erythroid-2 Levels Promote Epo-Independent Erythroid Maturation and Recapitulate the Hematopoietic Stem Cell and Common Myeloid Progenitor Expansion Observed in Polycythemia Vera Patients. Stem Cells Transl Med 2013;2(2):112–117.

25. Kaufmann KB, Gründer A, Hadlich T, et al. A novel murine model of myeloproliferative disorders generated by overexpression of the transcription factor NF-E2. J Exp Med 2012;209(1):35–50.

26. Scott LM, Tong W, Levine RL, et al. JAK2 exon 12 mutations in polycythemia vera and idiopathic erythrocytosis. N Engl J Med 2007;356(5):459–468.

27. Passamonti F, Elena C, Schnittger S, et al. Molecular and clinical features of the myeloproliferative neoplasm associated with JAK2 exon 12 mutations. Blood 2011;117(10):2813–2816.

28. Dawson MA, Bannister AJ, Göttgens B, et al. JAK2 phosphorylates histone H3Y41 and excludes HP1α from chromatin. Nature 2009;461(7265):819–822.

29. Plo I, Nakatake M, Malivert L, et al. JAK2 stimulates homologous recombination and genetic instability: potential implication in the heterogeneity of myeloproliferative disorders. Blood 2008;112(4):1402–1412.

30. Chen E, Ahn JS, Massie CE, et al. JAK2V617F promotes replication fork stalling with disease-restricted impairment of the intra-S checkpoint response. Proc Natl Acad Sci U S A 2014;111(42):15190–15195.

31. Ahn JS, Li J, Chen E, Kent DG, Park HJ, Green AR. JAK2V617F mediates resistance to DNA damage-induced apoptosis by modulating FOXO3A localization and Bcl-xL deamidation. Oncogene. 2016;35(17):2235-46

32. Anand S, Stedham F, Beer P, et al. Effects of the JAK2 mutation on the hematopoietic stem and progenitor compartment in human myeloproliferative neoplasms. Blood 2011;118(1):177–181.

33. Ishii T, Bruno E, Hoffman R, Xu M. Involvement of various hematopoietic-cell lineages by the JAK2V617F mutation in polycythemia vera. Blood 2006;108(9):3128–3134.

34. James C, Mazurier F, Dupont S, et al. The hematopoietic stem cell compartment of JAK2V617F-positive myeloproliferative disorders is a reflection of disease heterogeneity. Blood 2008;112(6):2429–2438.

35. Mullally A, Lane SW, Ball B, et al. Physiological Jak2V617F expression causes a lethal myeloproliferative neoplasm with differential effects on hematopoietic stem and progenitor cells. Cancer Cell 2010;17(6):584–596.

36. Li J, Kent DG, Godfrey AL, et al. JAK2V617F homozygosity drives a phenotypic switch in myeloproliferative neoplasms, but is insufficient to sustain disease. Blood 2014;123(20):3139–3151.

37. Pelt KV, Nollet F, Selleslag D, et al. The JAK2V617F mutation can occur in a hematopoietic stem cell that exhibits no proliferative advantage: a case of human allogeneic transplantation. Blood 2008;112(3):921–922.

38. Nielsen C, Bojesen SE, Nordestgaard BG, Kofoed KF, Birgens HS. JAK2V617F somatic mutation in the general population: myeloproliferative neoplasm development and progression rate. Haematologica 2014;99(9):1448–1455.

39. Sun J, Ramos A, Chapman B, et al. Clonal dynamics of native haematopoiesis. Nature 2014;514(7522):322–327.

40. Godfrey AL, Chen E, Pagano F, et al. JAK2V617F homozygosity arises commonly and recurrently in PV and ET, but PV is characterized by expansion of a dominant homozygous subclone. Blood 2012;120(13):2704–2707.

41. Vannucchi AM, Antonioli E, Guglielmelli P, et al. Clinical profile of homozygous JAK2 617V>F mutation in patients with polycythemia vera or essential thrombocythemia. Blood 2007;110(3):840–846.

42. Godfrey AL, Chen E, Pagano F, Silber Y, Campbell PJ, Green AR. Clonal analyses reveal associations of JAK2V617F homozygosity with hematologic features, age and gender in polycythemia vera and essential thrombocythemia. Haematologica 2013;98(5):718–721.

43. Tiedt R, Hao-Shen H, Sobas MA, et al. Ratio of mutant JAK2-V617F to wild-type Jak2 determines the MPD phenotypes in transgenic mice. Blood 2008;111(8):3931–

3940.

44. Saliba J, Hamidi S, Lenglet G, et al. Heterozygous and Homozygous JAK2V617F States Modeled by Induced Pluripotent Stem Cells from Myeloproliferative Neoplasm Patients. PLoS One. 2013;8(9):e74257

45. Chen E, Beer PA, Godfrey AL, et al. Distinct Clinical Phenotypes Associated with JAK2V617F Reflect Differential STAT1 Signaling. Cancer Cell 2010;18(5):524–535.

46. Pikman Y, Lee BH, Mercher T, et al. MPLW515L Is a Novel Somatic Activating Mutation in Myelofibrosis with Myeloid Metaplasia. PLoS Med 2006;3(7):e270.

47. Pardanani AD, Levine RL, Lasho T, et al. MPL515 mutations in myeloproliferative and other myeloid disorders: a study of 1182 patients. Blood 2006;108(10):3472–3476.

48. Morrell R, Langabeer SE, Smyth L, Perera M, Crotty G. Nonfamilial, MPL S505N-Mutated Essential Thrombocythaemia. Case Rep Hematol. 2013;2013:729327

49. Cabagnols X, Favale F, Pasquier F, et al. Presence of atypical thrombopoietin receptor (MPL) mutations in triple-negative essential thrombocythemia patients. Blood 2016;127(3):333–342.

50. Milosevic Feenstra JD, Nivarthi H, Gisslinger H, et al. Whole-exome sequencing identifies novel MPL and JAK2 mutations in triple-negative myeloproliferative neoplasms. Blood 2016;127(3):325–332.

51. Klampfl T, Gisslinger H, Harutyunyan AS, et al. Somatic Mutations of Calreticulin in Myeloproliferative Neoplasms. N Engl J Med 2013;369(25):2379–2390.

52. Nangalia J, Massie CE, Baxter EJ, et al. Somatic *CALR* Mutations in Myeloproliferative Neoplasms with Nonmutated *JAK2*. N Engl J Med 2013;369(25):2391–2405.

53. Arnaudeau S, Frieden M, Nakamura K, Castelbou C, Michalak M, Demaurex N. Calreticulin Differentially Modulates Calcium Uptake and Release in the Endoplasmic Reticulum and Mitochondria. J Biol Chem 2002;277(48):46696–46705.

54. Ellgaard L, Frickel E-M. Calnexin, calreticulin, and ERp57. Cell Biochem Biophys 2003;39(3):223–247.

55. Vannucchi AM, Rotunno G, Bartalucci N, et al. Calreticulin mutation-specific immunostaining in myeloproliferative neoplasms: pathogenetic insight and diagnostic value. Leukemia 2014;28(9):1811–1818.

56. Rampal R, Al-Shahrour F, Abdel-Wahab O, et al. Integrated genomic analysis illustrates the central role of JAK-STAT pathway activation in myeloproliferative neoplasm pathogenesis. Blood. 2014 May 29;123(22):e123-33.

57. Araki M, Yang Y, Masubuchi N, et al. Activation of the thrombopoietin receptor by mutant calreticulin in CALR-mutant myeloproliferative neoplasms. Blood 2016;127(10):1307–1316.

58. Chachoua I, Pecquet C, El-Khoury M, et al. Thrombopoietin receptor activation by

myeloproliferative neoplasm associated calreticulin mutants. Blood 2016;127(10):1325–1335.

59. Balligand T, Achouri Y, Pecquet C, et al. Pathologic activation of thrombopoietin receptor and JAK2-STAT5 pathway by frameshift mutants of mouse calreticulin. Leukemia 2016;30(8):1775–1778.

60. Marty C, Pecquet C, Nivarthi H, et al. Calreticulin mutants in mice induce an MPL-dependent thrombocytosis with frequent progression to myelofibrosis. Blood 2016;127(10):1317–1324.

61. Elf S, Abdelfattah NS, Chen E, et al. Mutant Calreticulin Requires Both Its Mutant C-terminus and the Thrombopoietin Receptor for Oncogenic Transformation. Cancer Discov 2016;6(4):368–381.

62. Li J, Prins D, Park HJ, et al. Mutant calreticulin knock-in mice develop thrombocytosis and myelofibrosis without a stem cell self-renewal advantage. Blood 2017;blood-2017-09-806356.

63. Tefferi A, Lasho TL, Tischer A, et al. The prognostic advantage of calreticulin mutations in myelofibrosis might be confined to type 1 or type 1-like CALR variants. Blood 2014;124(15):2465–2466.

64. on behalf of the Associazione Italiana per la Ricerca sul Cancro Gruppo Italiano Malattie Mieloproliferative (AGIMM), Guglielmelli P, Rotunno G, et al. Validation of the differential prognostic impact of type 1/type 1-like versus type 2/type 2-like CALR mutations in myelofibrosis. Blood Cancer J 2015;5(10):e360–e360.

65. Pietra D, Rumi E, Ferretti VV, et al. Differential clinical effects of different mutation subtypes in CALR-mutant myeloproliferative neoplasms. Leukemia. 2016;30(2):431-8.

66. Brecqueville M, Rey J, Bertucci F, et al. Mutation analysis of ASXL1, CBL, DNMT3A, IDH1, IDH2, JAK2, MPL, NF1, SF3B1, SUZ12, and TET2 in myeloproliferative neoplasms. Genes Chromosomes Cancer 2012;51(8):743–755.

67. Brecqueville M, Rey J, Devillier R, et al. Array comparative genomic hybridization and sequencing of 23 genes in 80 patients with myelofibrosis at chronic or acute phase. Haematologica 2014;99(1):37–45.

68. Tefferi A, Lasho TL, Guglielmelli P, et al. Targeted deep sequencing in polycythemia vera and essential thrombocythemia. Blood Adv 2016;1(1):21–30.

69. Tefferi A, Lasho TL, Finke CM, et al. Targeted deep sequencing in primary myelofibrosis. Blood Adv 2016;1(2):105–111.

70. Wang L, Swierczek SI, Drummond J, et al. Whole-exome sequencing of polycythemia vera revealed novel driver genes and somatic mutation shared by T cells and granulocytes. Leukemia 2014;28(4):935–938.

71. Velazquez L, Cheng AM, Fleming HE, et al. Cytokine Signaling and Hematopoietic Homeostasis Are Disrupted in Lnk-deficient Mice. J Exp Med 2002;195(12):1599–1611.

72. Lasho TL, Pardanani A, Tefferi A. LNK Mutations in JAK2 Mutation–Negative Erythrocytosis. N Engl J Med 2010;363(12):1189–1190.

73. Oh ST, Simonds EF, Jones C, et al. Novel mutations in the inhibitory adaptor protein LNK drive JAK-STAT signaling in patients with myeloproliferative neoplasms. Blood 2010;116(6):988–992.

74. Pardanani A, Lasho T, Finke C, Oh ST, Gotlib J, Tefferi A. LNK mutation studies in blast-phase myeloproliferative neoplasms, and in chronic-phase disease with TET2, IDH, JAK2 or MPL mutations. Leukemia 2010;24(10):1713–1718.

75. Toffalini F, Demoulin J-B. New insights into the mechanisms of hematopoietic cell transformation by activated receptor tyrosine kinases. Blood 2010;116(14):2429–2437.

76. Grand FH, Hidalgo-Curtis CE, Ernst T, et al. Frequent CBL mutations associated with 11q acquired uniparental disomy in myeloproliferative neoplasms. Blood 2009;113(24):6182–6192.

77. Schwaab J, Ernst T, Erben P, et al. Activating CBL mutations are associated with a distinct MDS/MPN phenotype. Ann Hematol 2012;91(11):1713–1720.

78. Suessmuth Y, Elliott J, Percy MJ, et al. A new polycythaemia vera-associated SOCS3 SH2 mutant (SOCS3F136L) cannot regulate erythropoietin responses. Br J Haematol 2009;147(4):450–458.

79. Jost E, do Ó N, Dahl E, et al. Epigenetic alterations complement mutation of JAK2 tyrosine kinase in patients with BCR/ABL-negative myeloproliferative disorders. Leukemia 2007;21(3):505–510.

80. Janssen JW, Steenvoorden AC, Lyons J, et al. RAS gene mutations in acute and chronic myelocytic leukemias, chronic myeloproliferative disorders, and myelodysplastic syndromes. Proc Natl Acad Sci U S A 1987;84(24):9228–9232.

81. Wang J, Kong G, Liu Y, et al. NrasG12D/+ promotes leukemogenesis by aberrantly regulating hematopoietic stem cell functions. Blood 2013;121(26):5203–5207.

82. Kratz CP. The mutational spectrum of PTPN11 in juvenile myelomonocytic leukemia and Noonan syndrome/myeloproliferative disease. Blood 2005;106(6):2183–2185.

83. Delhommeau F, Dupont S, Valle VD, et al. Mutation in TET2 in Myeloid Cancers. N Engl J Med 2009;360(22):2289–2301.

84. Tefferi A, Pardanani A, Lim K-H, et al. TET2 mutations and their clinical correlates in polycythemia vera, essential thrombocythemia and myelofibrosis. Leukemia 2009;23(5):905–911.

85. Ko M, Huang Y, Jankowska AM, et al. Impaired hydroxylation of 5-methylcytosine in myeloid cancers with mutant TET2. Nature 2010;468(7325):839–843.

86. Ko M, Bandukwala HS, An J, et al. Ten-Eleven-Translocation 2 (TET2) negatively regulates homeostasis and differentiation of hematopoietic stem cells in mice. Proc Natl Acad Sci 2011;108(35):14566–14571.

87. Moran-Crusio K, Reavie L, Shih A, et al. Tet2 loss leads to increased hematopoietic stem cell self-renewal and myeloid transformation. Cancer Cell 2011;20(1):11–24.

88. Bocker MT, Tuorto F, Raddatz G, et al. Hydroxylation of 5-methylcytosine by TET2 maintains the active state of the mammalian HOXA cluster. Nat Commun 2012;3818.

89. Stegelmann F, Bullinger L, Schlenk RF, et al. DNMT3A mutations in myeloproliferative neoplasms. Leukemia 2011;25(7):1217–1219.

90. Challen GA, Sun D, Jeong M, et al. Dnmt3a is essential for hematopoietic stem cell differentiation. Nat Genet 2011;44(1):23–31.

91. Mayle A, Yang L, Rodriguez B, et al. Dnmt3a loss predisposes murine hematopoietic stem cells to malignant transformation. Blood 2015;125(4):629–638.

92. Green A, Beer P. Somatic Mutations of IDH1 and IDH2 in the Leukemic Transformation of Myeloproliferative Neoplasms. N Engl J Med 2010;362(4):369–370.

93. Dang L, White DW, Gross S, et al. Cancer-associated IDH1 mutations produce 2-hydroxyglutarate. Nature 2009;462(7274):739–744.

94. Figueroa ME, Abdel-Wahab O, Lu C, et al. Leukemic IDH1 and IDH2 Mutations Result in a Hypermethylation Phenotype, Disrupt TET2 Function, and Impair Hematopoietic Differentiation. Cancer Cell 2010;18(6):553–567.

95. Lundberg P, Karow A, Nienhold R, et al. Clonal evolution and clinical correlates of somatic mutations in myeloproliferative neoplasms. Blood 2014;123(14):2220–2228.

96. Tefferi A, Jimma T, Sulai NH, et al. IDH mutations in primary myelofibrosis predict leukemic transformation and shortened survival: clinical evidence for leukemogenic collaboration with JAK2V617F. Leukemia 2012;26(3):475–480.

97. Ernst T, Chase AJ, Score J, et al. Inactivating mutations of the histone methyltransferase gene EZH2 in myeloid disorders. Nat Genet 2010;42(8):722–726.

98. Guglielmelli P, Biamonte F, Score J, et al. EZH2 mutational status predicts poor survival in myelofibrosis. Blood 2011;118(19):5227–5234.

99. Muto T, Sashida G, Oshima M, et al. Concurrent loss of Ezh2 and Tet2 cooperates in the pathogenesis of myelodysplastic disorders. J Exp Med 2013;210(12):2627–2639.

100. Khan SN, Jankowska AM, Mahfouz R, et al. Multiple mechanisms deregulate EZH2 and histone H3 lysine 27 epigenetic changes in myeloid malignancies. Leukemia 2013;27(6):1301–1309.

101. Carbuccia N, Murati A, Trouplin V, et al. Mutations of ASXL1 gene in myeloproliferative neoplasms. Leukemia 2009;23(11):2183–2186.

102. Abdel-Wahab O, Adli M, LaFave LM, et al. ASXL1 Mutations Promote Myeloid Transformation Through Loss of PRC2-Mediated Gene Repression. Cancer Cell 2012;22(2):180–193.

103. Lasho TL, Jimma T, Finke CM, et al. SRSF2 mutations in primary myelofibrosis:

significant clustering with IDH mutations and independent association with inferior overall and leukemia-free survival. Blood 2012;120(20):4168–4171.

104. Zhang S-J, Rampal R, Manshouri T, et al. Genetic analysis of patients with leukemic transformation of myeloproliferative neoplasms shows recurrent SRSF2 mutations that are associated with adverse outcome. Blood 2012;119(19):4480–4485.

105. Tefferi A, Finke CM, Lasho TL, et al. U2AF1 mutations in primary myelofibrosis are strongly associated with anemia and thrombocytopenia despite clustering with JAK2V617F and normal karyotype. Leukemia 2014;28(2):431–433.

106. Kim E, Ilagan JO, Liang Y, et al. SRSF2 Mutations Contribute to Myelodysplasia by Mutant-Specific Effects on Exon Recognition. Cancer Cell 2015;27(5):617–630.

107. Shirai CL, Ley JN, White BS, et al. Mutant U2AF1 Expression Alters Hematopoiesis and Pre-mRNA Splicing In Vivo. Cancer Cell 2015;27(5):631–643.

108. Mupo A, Seiler M, Sathiaseelan V, et al. Hemopoietic-specific *Sf3b1*-K700E knock-in mice display the splicing defect seen in human MDS but develop anemia without ring sideroblasts. Leukemia 2017;31(3):720–727.

109. Quesada V, Conde L, Villamor N, et al. Exome sequencing identifies recurrent mutations of the splicing factor *SF3B1* gene in chronic lymphocytic leukemia. Nat Genet 2012;44(1):47–52.

110. Papaemmanuil E, Cazzola M, Boultwood J, et al. Somatic SF3B1 Mutation in Myelodysplasia with Ring Sideroblasts. N Engl J Med 2011;365(15):1384–1395.

111. Darman RB, Seiler M, Agrawal AA, et al. Cancer-Associated SF3B1 Hotspot Mutations Induce Cryptic 3′ Splice Site Selection through Use of a Different Branch Point. Cell Rep 2015;13(5):1033–1045.

112. Dolatshad H, Pellagatti A, Fernandez-Mercado M, et al. Disruption of SF3B1 results in deregulated expression and splicing of key genes and pathways in myelodysplastic syndrome hematopoietic stem and progenitor cells. Leukemia. 2015 May;29(5):1092-103.

113. Jutzi JS, Bogeska R, Nikoloski G, et al. MPN patients harbor recurrent truncating mutations in transcription factor NF-E2. J Exp Med 2013;210(5):1003–1019.

114. Gasiorek JJ, Blank V. Regulation and function of the NFE2 transcription factor in hematopoietic and non-hematopoietic cells. Cell Mol Life Sci 2015;72(12):2323–2335.

115. Kapralova K, Lanikova L, Lorenzo F, et al. RUNX1 and NF-E2 upregulation is not specific for MPNs, but is seen in polycythemic disorders with augmented HIF signaling. Blood 2014;123(3):391–394.

116. Lasho TL, Mudireddy M, Finke CM, et al. Targeted next-generation sequencing in blast phase myeloproliferative neoplasms. Blood Adv 2018;2(4):370–380.

117. Alford KA, Reinhardt K, Garnett C, et al. Analysis of GATA1 mutations in Down syndrome transient myeloproliferative disorder and myeloid leukemia. Blood 2011;118(8):2222–2238.

118. Jäger R, Gisslinger H, Passamonti F, et al. Deletions of the transcription factor *Ikaros* in myeloproliferative neoplasms. Leukemia 2010;24(7):1290–1298.

119. Kon A, Shih L-Y, Minamino M, et al. Recurrent mutations in multiple components of the cohesin complex in myeloid neoplasms. Nat Genet 2013;45(10):1232–1237.

120. Todd MAM, Ivanochko D, Picketts DJ. PHF6 Degrees of Separation: The Multifaceted Roles of a Chromatin Adaptor Protein. Genes 2015;6(2):325–352.

121. Prick J, de Haan G, Green AR, Kent DG. Clonal heterogeneity as a driver of disease variability in the evolution of myeloproliferative neoplasms. Exp Hematol 2014;42(10):841–851.

122. Ortmann CA, Kent DG, Nangalia J, et al. Effect of Mutation Order on Myeloproliferative Neoplasms. N Engl J Med 2015;372(7):601–612.

123. Nangalia J, Nice FL, Wedge DC, et al. DNMT3A mutations occur early or late in patients with myeloproliferative neoplasms and mutation order influences phenotype. Haematologica 2015;100(11):e438–e442.

124. Inda M-M, Bonavia R, Mukasa A, et al. Tumor heterogeneity is an active process maintained by a mutant EGFR-induced cytokine circuit in glioblastoma. Genes Dev 2010;24(16):1731–1745.

125. Klampfl T, Harutyunyan A, Berg T, et al. Genome integrity of myeloproliferative neoplasms in chronic phase and during disease progression. Blood 2011;118(1):167–176.

126. Rumi E, Harutyunyan A, Elena C, et al. Identification of genomic aberrations associated with disease transformation by means of high-resolution SNP array analysis in patients with myeloproliferative neoplasm. Am J Hematol 2011;86(12):974–979.

127. Shen W, Paxton CN, Szankasi P, et al. Detection of genome-wide copy number variants in myeloid malignancies using next-generation sequencing. J Clin Pathol 2018;71(4):372–378.

128. Stegelmann F, Bullinger L, Griesshammer M, et al. High-resolution single-nucleotide polymorphism array-profiling in myeloproliferative neoplasms identifies novel genomic aberrations. Haematologica 2010;95(4):666–669.

129. Harutyunyan A, Gisslinger B, Klampfl T, et al. Rare germline variants in regions of loss of heterozygosity may influence clinical course of hematological malignancies. Leukemia 2011;25(11):1782–1784.

130. Ulirsch JC, Nandakumar SK, Wang L, et al. Systematic Functional Dissection of Common Genetic Variation Affecting Red Blood Cell Traits. Cell 2016;165(6):1530–1545.

131. Teofili L, Giona F, Torti L, et al. Hereditary thrombocytosis caused by MPLSer505Asn is associated with a high thrombotic risk, splenomegaly and progression to bone marrow fibrosis. Haematologica 2010;95(1):65–70.

132. Posthuma HLA, Skoda RC, Jacob FA, Maas APC van der, Valk PJM, Posthuma EFM. Hereditary thrombocytosis not as innocent as thought? Development into acute

leukemia and myelofibrosis. Blood 2010;116(17):3375–3376.

133. Ma W, Kantarjian H, Zhang X, et al. Mutation Profile of JAK2 Transcripts in Patients with Chronic Myeloproliferative Neoplasias. J Mol Diagn 2009;11(1):49–53.

134. Mead AJ, Rugless MJ, Jacobsen SEW, Schuh A. Germline JAK2 Mutation in a Family with Hereditary Thrombocytosis. N Engl J Med. 2012;366(10):967-9.

135. Kapralova K, Horvathova M, Pecquet C, et al. Cooperation of germline JAK2 mutations E846D and R1063H in hereditary erythrocytosis with megakaryocytic atypia. Blood 2016;blood-2016-02-698951.

136. Jones AV, Cross NCP. Inherited predisposition to myeloproliferative neoplasms. Ther Adv Hematol 2013;4(4):237–253.

137. Giambruno R, Krendl C, Stukalov A, et al. Germline RBBP6 Mutations In Myeloproliferative Neoplasms. Blood 2013;122(21):267–267.

138. Saliba J, Saint-Martin C, Di Stefano A, et al. Germline duplication of ATG2B and GSKIP predisposes to familial myeloid malignancies. Nat Genet. 2015 Oct;47(10):1131-40.

139. Oddsson A, Kristinsson SY, Helgason H, et al. The germline sequence variant rs2736100_C in TERT associates with myeloproliferative neoplasms. Leukemia 2014;28(6):1371–1374.

140. Tapper W, Jones AV, Kralovics R, et al. Genetic variation at MECOM, TERT, JAK2 and HBS1L-MYB predisposes to myeloproliferative neoplasms. Nat Commun 2015;66691.

141. Jones AV, Chase A, Silver RT, et al. JAK2 haplotype is a major risk factor for the development of myeloproliferative neoplasms. Nat Genet 2009;41(4):446–449.

142. Olcaydu D, Harutyunyan A, Jäger R, et al. A common JAK2 haplotype confers susceptibility to myeloproliferative neoplasms. Nat Genet 2009;41(4):450–454.

143. Olcaydu D, Skoda RC, Looser R, et al. The “GGCC” haplotype of JAK2 confers susceptibility to JAK2 exon 12 mutation-positive polycythemia vera. Leukemia 2009;23(10):1924–1926.

144. Alvarez-Larrán A, Angona A, Martínez-Avilés L, Bellosillo B, Besses C. Influence of JAK2 46/1 haplotype in the natural evolution of JAK2V617F allele burden in patients with myeloproliferative neoplasms. Leuk Res 2012;36(3):324–326.

145. McKerrell T, Park N, Chi J, et al. JAK2 V617F hematopoietic clones are present several years prior to MPN diagnosis and follow different expansion kinetics. Blood Adv 2017;1(14):968–971.

146. Jones AV, Campbell PJ, Beer PA, et al. The JAK2 46/1 haplotype predisposes to MPL-mutated myeloproliferative neoplasms. Blood 2010;115(22):4517–4523.

147. Pardanani A, Fridley BL, Lasho TL, Gilliland DG, Tefferi A. Host genetic variation contributes to phenotypic diversity in myeloproliferative disorders. Blood 2008;111(5):2785–2789.

148. Varricchio L, Masselli E, Alfani E, et al. The dominant negative β isoform of the glucocorticoid receptor is uniquely expressed in erythroid cells expanded from polycythemia vera patients. Blood 2011;118(2):425–436.

149. Poletto V, Rosti V, Villani L, et al. A3669G polymorphism of glucocorticoid receptor is a susceptibility allele for primary myelofibrosis and contributes to phenotypic diversity and blast transformation. Blood 2012;120(15):3112–3117.

150. Costache RM, Bănescu C, Popp RA, Pop IV, Trifa AP. The glucocorticoid receptor A3669G SNP is not associated with polycythemia vera, essential thrombocythemia or primary myelofibrosis. Leuk Lymphoma 2016;57(1):209–211.

151. Patterson-Fortin J, Moliterno AR. Molecular Pathogenesis of Myeloproliferative Neoplasms: Influence of Age and Gender. Curr Hematol Malig Rep 2017;12(5):424–431.

152. Sidon P, Housni HE, Dessars B, Heimann P. The *JAK2*V617F mutation is detectable at very low level in peripheral blood of healthy donors. Leukemia 2006;20(9):1622.

153. Xu X, Zhang Q, Luo J, et al. JAK2V617F: prevalence in a large Chinese hospital population. Blood 2007;109(1):339–342.

154. Genovese G, Kähler AK, Handsaker RE, et al. Clonal Hematopoiesis and Blood-Cancer Risk Inferred from Blood DNA Sequence. N Engl J Med. 2014;371(26):2477-87.

155. Jaiswal S, Fontanillas P, Flannick J, et al. Age-Related Clonal Hematopoiesis Associated with Adverse Outcomes. N Engl J Med. 2014;371(26):2488-98..

156. McKerrell T, Park N, Moreno T, et al. Leukemia-Associated Somatic Mutations Drive Distinct Patterns of Age-Related Clonal Hemopoiesis. Cell Rep 2015;10(8):1239–1245.

157. Xie M, Lu C, Wang J, et al. Age-related mutations associated with clonal hematopoietic expansion and malignancies. Nat Med. 2014;20(12):1472-8.

158. Young AL, Challen GA, Birmann BM, Druley TE. Clonal haematopoiesis harbouring AML-associated mutations is ubiquitous in healthy adults. Nat Commun 2016;712484.

159. Hinds DA, Barnholt KE, Mesa RA, et al. Germ line variants predispose to both JAK2 V617F clonal hematopoiesis and myeloproliferative neoplasms. Blood 2016;128(8):1121–1128.

160. Nielsen C, Birgens HS, Nordestgaard BG, Bojesen SE. Diagnostic value of *JAK2* V617F somatic mutation for myeloproliferative cancer in 49 488 individuals from the general population. Br J Haematol 2013;160(1):70–79.

161. McKerrell T, Vassiliou GS. Aging as a driver of leukemogenesis. Sci Transl Med 2015;7(306):306fs38-306fs38.

162. Kim Y-W, Koo B-K, Jeong H-W, et al. Defective Notch activation in microenvironment leads to myeloproliferative disease. Blood 2008;112(12):4628–4638.

163. Walkley CR, Olsen GH, Dworkin S, et al. A Microenvironment-Induced Myeloproliferative Syndrome Caused by Retinoic Acid Receptor γ Deficiency. Cell 2007;129(6):1097–1110.

164. Fleischman AG, Aichberger KJ, Luty SB, et al. TNF facilitates clonal expansion of JAK2V617F positive cells in myeloproliferative neoplasms. Blood 2011;118(24):6392–6398.

165. Manshouri T, Estrov Z, Quintas-Cardama A, et al. Bone Marrow Stroma-Secreted Cytokines Protect JAK2V617F-Mutated Cells from the Effects of a JAK2 Inhibitor. Cancer Res 2011;71(11):3831–3840.

166. Hasselbalch HC. The role of cytokines in the initiation and progression of myelofibrosis. Cytokine Growth Factor Rev 2013;24(2):133–145.

167. Kleppe M, Kwak M, Koppikar P, et al. JAK-STAT Pathway Activation in Malignant and Nonmalignant Cells Contributes to MPN Pathogenesis and Therapeutic Response. Cancer Discov 2015;5(3):316–331.

168. Kagoya Y, Yoshimi A, Tsuruta-Kishino T, et al. JAK2V617F+ myeloproliferative neoplasm clones evoke paracrine DNA damage to adjacent normal cells through secretion of lipocalin-2. Blood 2014;124(19):2996–3006.

169. Arranz L, Sánchez-Aguilera A, Martín-Pérez D, et al. Neuropathy of haematopoietic stem cell niche is essential for myeloproliferative neoplasms. Nature. 2014;512(7512):78-81.

170. Schepers K, Pietras EM, Reynaud D, et al. Myeloproliferative Neoplasia Remodels the Endosteal Bone Marrow Niche into a Self-Reinforcing Leukemic Niche. Cell Stem Cell 2013;13(3):285–299.

171. Andersen M, Sajid Z, Pedersen RK, et al. Mathematical modelling as a proof of concept for MPNs as a human inflammation model for cancer development. PloS One 2017;12(8):e0183620.

172. Hasselbalch HC, Bjørn ME. MPNs as Inflammatory Diseases: The Evidence, Consequences, and Perspectives. Mediators Inflamm. 2015;2015:102476.

173. Hoermann G, Greiner G, Valent P. Cytokine Regulation of Microenvironmental Cells in Myeloproliferative Neoplasms. Mediators Inflamm 2015;2015:869242.

174. Zhang J, Fleischman AG, Wodarz D, Komarova NL. Determining the role of inflammation in the selection of JAK2 mutant cells in myeloproliferative neoplasms. J Theor Biol 2017;42543–52.

175. Panteli KE, Hatzimichael EC, Bouranta PK, et al. Serum interleukin (IL)-1, IL-2, sIL-2Ra, IL-6 and thrombopoietin levels in patients with chronic myeloproliferative diseases. Br J Haematol 2005;130(5):709–715.

176. Ho C-L, Lasho TL, Butterfield JH, Tefferi A. Global cytokine analysis in myeloproliferative disorders. Leuk Res 2007;31(10):1389–1392.

177. Tefferi A, Vaidya R, Caramazza D, Finke C, Lasho T, Pardanani A. Circulating Interleukin (IL)-8, IL-2R, IL-12, and IL-15 Levels Are Independently Prognostic in

Primary Myelofibrosis: A Comprehensive Cytokine Profiling Study. J Clin Oncol 2011;29(10):1356–1363.

178. Pourcelot E, Trocme C, Mondet J, Bailly S, Toussaint B, Mossuz P. Cytokine profiles in polycythemia vera and essential thrombocythemia patients: Clinical implications. Exp Hematol 2014;42(5):360–368.

179. Boissinot M, Cleyrat C, Vilaine M, Jacques Y, Corre I, Hermouet S. Anti-inflammatory cytokines hepatocyte growth factor and interleukin-11 are over-expressed in Polycythemia vera and contribute to the growth of clonal erythroblasts independently of JAK2V617F. Oncogene 2011;30(8):990–1001.

180. Salim JP, Goette NP, Lev PR, et al. Dysregulation of stromal derived factor 1/CXCR4 axis in the megakaryocytic lineage in essential thrombocythemia. Br J Haematol 2009;144(1):69–77.

181. Harrison CN, Campbell PJ, Buck G, et al. Hydroxyurea compared with anagrelide in high-risk essential thrombocythemia. N Engl J Med 2005;353(1):33–45.

182. Godfrey AL, Campbell PJ, MacLean C, et al. Hydroxycarbamide Plus Aspirin Versus Aspirin Alone in Patients With Essential Thrombocythemia Age 40 to 59 Years Without High-Risk Features. J Clin Oncol 2018;JCO.2018.78.8414.

183. Harrison CN, Butt N, Campbell P, et al. Modification of British Committee for Standards in Haematology diagnostic criteria for essential thrombocythaemia. Br J Haematol. 2014;167(3):421-3.

184. Andersen CL, McMullin MF, Ejerblad E, et al. A phase II study of vorinostat (MK-0683) in patients with polycythaemia vera and essential thrombocythaemia. Br J Haematol 2013;162(4):498–508.

185. Andersen CL, Mortensen NB, Klausen TW, Vestergaard H, Bjerrum OW, Hasselbalch HC. A phase II study of vorinostat (MK-0683) in patients with primary myelofibrosis and post-polycythemia vera myelofibrosis. Haematologica 2014;99(1):e5–e7.

186. McMullin MF, Bareford D, Campbell P, et al. Guidelines for the diagnosis, investigation and management of polycythaemia/erythrocytosis. Br J Haematol 2005;130(2):174–195.

187. McMullin MF, Reilly JT, Campbell P, et al. Amendment to the guideline for diagnosis and investigation of polycythaemia/erythrocytosis. Br J Haematol 2007;138(6):821–822.

188. Reilly JT, McMullin MF, Beer PA, et al. Guideline for the diagnosis and management of myelofibrosis. Br J Haematol 2012;158(4):453–471.

189. Papaemmanuil E, Gerstung M, Malcovati L, et al. Clinical and biological implications of driver mutations in myelodysplastic syndromes. Blood 2013;122(22):3616–3627.

190. Li H, Durbin R. Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics 2009;25(14):1754–1760.

191. Jones AV, Ward D, Lyon M, et al. Evaluation of methods to detect CALR mutations

in myeloproliferative neoplasms. Leuk Res 2015;39(1):82–87.

192. Campbell PJ, Scott LM, Buck G, et al. Definition of subtypes of essential thrombocythaemia and relation to polycythaemia vera based on JAK2 V617F mutation status: a prospective study. The Lancet 2005;366(9501):1945–1953.

193. Beer PA, Campbell PJ, Scott LM, et al. MPL mutations in myeloproliferative disorders: analysis of the PT-1 cohort. Blood 2008;112(1):141–149.

194. Jones D, Raine KM, Davies H, et al. cgpCaVEManWrapper: Simple Execution of CaVEMan in Order to Detect Somatic Single Nucleotide Variants in NGS Data. Curr Protoc Bioinforma 2016;5615.10.1-15.10.18.

195. Ye K, Schulz MH, Long Q, Apweiler R, Ning Z. Pindel: a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. Bioinformatics 2009;25(21):2865–2871.

196. Papaemmanuil E, Cazzola M, Boultwood J, et al. Somatic SF3B1 mutation in myelodysplasia with ring sideroblasts. N Engl J Med 2011;365(15):1384–1395.

197. 1000 Genomes Project Consortium, Abecasis GR, Auton A, et al. An integrated map of genetic variation from 1,092 human genomes. Nature 2012;491(7422):56–65.

198. Exome Variant Server. http://evs.gs.washington.edu/EVS/ (accessed April 15, 2015).

199. Smigielski EM, Sirotkin K, Ward M, Sherry ST. dbSNP: a database of single nucleotide polymorphisms. Nucleic Acids Res 2000;28(1):352–355.

200. Lek M, Karczewski KJ, Minikel EV, et al. Analysis of protein-coding genetic variation in 60,706 humans. Nature 2016;536(7616):285–291.

201. Forbes SA, Tang G, Bindal N, et al. COSMIC (the Catalogue of Somatic Mutations in Cancer): a resource to investigate acquired mutations in human cancer. Nucleic Acids Res 2010;38(Database issue):D652-657.

202. Forbes SA, Bindal N, Bamford S, et al. COSMIC: mining complete cancer genomes in the Catalogue of Somatic Mutations in Cancer. Nucleic Acids Res 2010;gkq929.

203. Cosmic. COSMIC - Catalogue of Somatic Mutations in Cancer. https://cancer.sanger.ac.uk/cosmic (accessed April 15, 2015).

204. Li H, Handsaker B, Wysoker A, et al. The Sequence Alignment/Map format and SAMtools. Bioinforma Oxf Engl 2009;25(16):2078–2079.

205. Van der Harst P, Zhang W, Mateo Leach I, et al. Seventy-five genetic loci influencing the human red blood cell. Nature 2012;492(7429):369–375.

206. Gieger C, Radhakrishnan A, Cvejic A, et al. New gene functions in megakaryopoiesis and platelet formation. Nature 2011;480(7376):201–208.

207. Shameer K, Denny JC, Ding K, et al. A Genome- and Phenome-Wide Association Study to Identify Genetic Variants Influencing Platelet Count and Volume and their Pleiotropic Effects. Hum Genet;133(1):.

208. Kamatani Y, Matsuda K, Okada Y, et al. Genome-wide association study of

hematological and biochemical traits in a Japanese population. Nat Genet 2010;42(3):210–215.

209. Ganesh SK, Zakai NA, van Rooij FJA, et al. Multiple loci influence erythrocyte phenotypes in the CHARGE Consortium. Nat Genet 2009;41(11):1191–1198.

210. Ensembl Genome Browser. http://dec2013.archive.ensembl.org/index.html (accessed October 15, 2015).

211. Bartlett M, Cussens J. Integer Linear Programming for the Bayesian network structure learning problem. Artif Intell 2017;244258–271.

212. Wielemaker J, Schrijvers T, Triska M, Lager T. SWI-Prolog. Theory Pract Log Program 2012;12(1–2):67–96.

213. Nangalia J, Nice FL, Wedge DC, et al. DNMT3A mutations occur early or late in patients with myeloproliferative neoplasms and mutation order influences phenotype. Haematologica 2015;100(11):e438–e442.

214. Papaemmanuil E, Gerstung M, Bullinger L, et al. Genomic Classification and Prognosis in Acute Myeloid Leukemia. N Engl J Med 2016;374(23):2209–2221.

215. Gerstung M, Papaemmanuil E, Martincorena I, et al. Precision oncology for acute myeloid leukemia using a knowledge bank approach. Nat Genet 2017;49(3):332–340.

216. Steyerberg EW, Harrell FE, Borsboom GJJM, Eijkemans MJC, Vergouwe Y, Habbema JDF. Internal validation of predictive models: Efficiency of some procedures for logistic regression analysis. J Clin Epidemiol 2001;54(8):774–781.

217. Uno H, Cai T, Pencina MJ, D'Agostino RB, Wei LJ. On the C-statistics for evaluating overall adequacy of risk prediction procedures with censored survival data. Stat Med 2011;n/a-n/a.

218. Brier GW. Verification of forecasts expressed in terms of probability. Mon Weather Rev 1950;78(1):1–3.

219. Ferro CAT. Comparing Probabilistic Forecasting Systems with the Brier Score. Weather Forecast 2007;22(5):1076–1088.

220. Akbari MR, Lepage P, Rosen B, et al. PPM1D Mutations in Circulating White Blood Cells and the Risk for Ovarian Cancer. J Natl Cancer Inst 2013;djt323.

221. Ruark E, Snape K, Humburg P, et al. Mosaic PPM1D mutations are associated with predisposition to breast and ovarian cancer. Nature 2013;493(7432):406–410.

222. Dudgeon C, Shreeram S, Tanoue K, et al. Genetic variants and mutations of PPM1D control the response to DNA damage. Cell Cycle 2013;12(16):2656–2664.

223. Kleiblova P, Shaltiel IA, Benada J, et al. Gain-of-function mutations of PPM1D/Wip1 impair the p53-dependent G1 checkpoint. J Cell Biol 2013;201(4):511–521.

224. Chen C, Liu Y, Rappaport AR, et al. MLL3 Is a Haploinsufficient 7q Tumor Suppressor in Acute Myeloid Leukemia. Cancer Cell 2014;25(5):652–665.

225. Steensma DP, Higgs DR, Fisher CA, Gibbons RJ. Acquired somatic ATRX mutations

in myelodysplastic syndrome associated with α thalassemia (ATMDS) convey a more severe hematologic phenotype than germline ATRX mutations. Blood 2004;103(6):2019–2026.

226. Horton SJ, Giotopoulos G, Yun H, et al. Early loss of Crebbp confers malignant stem cell properties on lymphoid progenitors. Nat Cell Biol 2017;19(9):1093–1104.

227. Li C, Franklin JL, Graves-Deal R, Jerome WG, Cao Z, Coffey RJ. Myristoylated Naked2 escorts transforming growth factor alpha to the basolateral plasma membrane of polarized epithelial cells. Proc Natl Acad Sci U S A 2004;101(15):5571–5576.

228. Li X-X, Zhou J-D, Zhang T-J, et al. Epigenetic dysregulation of NKD2 is a valuable predictor assessing treatment outcome in acute myeloid leukemia. J Cancer 2017;8(3):460–468.

229. Soler G, Bernal-Vicente A, Antón AI, et al. The JAK2 46/1 haplotype does not predispose to CALR-mutated myeloproliferative neoplasms. Ann Hematol 2014;94(5):789–794.

230. Koren A, Handsaker RE, Kamitaki N, et al. Genetic Variation in Human DNA Replication Timing. Cell 2014;159(5):1015–1026.

231. Yamamoto K. Stochastic model of Zipf's law and the universality of the power-law exponent. Phys Rev E Stat Nonlin Soft Matter Phys 2014;89(4):42115.

232. Khoury ME, Vertenoeil G, Marty C, et al. Calreticulin Mutants Induce an Early Clonal Dominance and a Megakaryocytic Phenotype through the Activation of MPL/JAK2 Pathway in Human Primary Cells. Blood 2016;128(22):1959–1959.

233. Vannucchi AM, Lasho TL, Guglielmelli P, et al. Mutations and prognosis in primary myelofibrosis. Leukemia 2013;27(9):1861–1869.

234. Guglielmelli P, Lasho TL, Rotunno G, et al. The number of prognostically detrimental mutations and prognosis in primary myelofibrosis: an international study of 797 patients. Leukemia 2014;28(9):1804–1810.

235. The Collected Mathematical Pappers of Arthur Cayley, Sc.d., F.r.s. CUP Archive.; 588 p.

236. Passamonti F, Cervantes F, Vannucchi AM, et al. A dynamic prognostic model to predict survival in primary myelofibrosis: a study by the IWG-MRT (International Working Group for Myeloproliferative Neoplasms Research and Treatment). Blood 2010;115(9):1703–1708.

237. Cervantes F, Dupriez B, Pereira A, et al. New prognostic scoring system for primary myelofibrosis based on a study of the International Working Group for Myelofibrosis Research and Treatment. Blood 2009;113(13):2895–2901.

238. Gangat N, Caramazza D, Vaidya R, et al. DIPSS Plus: A Refined Dynamic International Prognostic Scoring System for Primary Myelofibrosis That Incorporates Prognostic Information From Karyotype, Platelet Count, and Transfusion Status. J Clin Oncol 2011;29(4):392–397.

239. Vannucchi AM, Guglielmelli P, Rotunno G, et al. Mutation-Enhanced International

Prognostic Scoring System (MIPSS) for Primary Myelofibrosis: An AGIMM & IWG-MRT Project. Blood 2014;124(21):405–405.

240. Tefferi A, Guglielmelli P, Lasho TL, et al. CALR and ASXL1 mutations-based molecular prognostication in primary myelofibrosis: an international study of 570 patients. Leukemia 2014;28(7):1494–1500.

241. Tefferi A, Guglielmelli P, Nicolosi M, et al. GIPSS: genetically inspired prognostic scoring system for primary myelofibrosis. Leukemia 2018;32(7):1631–1642.

242. Passamonti F, Thiele J, Girodon F, et al. A prognostic model to predict survival in 867 World Health Organization-defined essential thrombocythemia at diagnosis: a study by the International Working Group on Myelofibrosis Research and Treatment. Blood 2012;120(6):1197–1201.

243. Schwarz G. Estimating the Dimension of a Model. Ann Stat 1978;6(2):461–464.

244. Harrell Jr FE. Regression modeling strategies: with applications to linear models, logistic and ordinal regression, and survival analysis. Springer-Verlag New York, 2001.

245. Abelson S, Collord G, Ng SWK, et al. Prediction of acute myeloid leukaemia risk in healthy individuals. Nature 2018;559(7714):400–404.

246. Hoermann G, Greiner G, Valent P. Cytokine Regulation of Microenvironmental Cells in Myeloproliferative Neoplasms. Mediators Inflamm 2015;20151–17.

247. Tefferi A, Vaidya R, Caramazza D, Finke C, Lasho T, Pardanani A. Circulating Interleukin (IL)-8, IL-2R, IL-12, and IL-15 Levels Are Independently Prognostic in Primary Myelofibrosis: A Comprehensive Cytokine Profiling Study. J Clin Oncol 2011;29(10):1356–1363.

248. Lekovic D, Gotic M, Skoda R, et al. Bone marrow microvessel density and plasma angiogenic factors in myeloproliferative neoplasms: clinicopathological and molecular correlations. Ann Hematol 2017;96(3):393–404.

249. Yoda A, Adelmant G, Tamburini J, et al. Mutations in G protein β subunits promote transformation and kinase inhibitor resistance. Nat Med 2015;21(1):71–75.

250. Ostrander EL, Koh WK, Mallaney C, et al. The GNASR201C mutation associated with clonal hematopoiesis supports transplantable hematopoietic stem cell activity. Exp Hematol 2018;5714–20.

251. Zhang L, Chen LH, Wan H, et al. Exome sequencing identifies somatic gain-of-function PPM1D mutations in brainstem gliomas. Nat Genet 2014;46(7):726–730.

252. Kahn JD, Miller PG, Silver AJ, et al. PPM1D-truncating mutations confer resistance to chemotherapy and sensitivity to PPM1D inhibition in hematopoietic cells. Blood 2018;132(11):1095–1105.

253. Hasselbalch HC. A role of NF-E2 in chronic inflammation and clonal evolution in essential thrombocythemia, polycythemia vera and myelofibrosis? Leuk Res 2014;38(2):263–266.

254. Wang L, Swierczek SI, Lanikova L, et al. The relationship of JAK2V617F and acquired UPD at chromosome 9p in polycythemia vera. Leukemia 2014;28(4):938–941.

255. Barrett JC, Hansoul S, Nicolae DL, et al. Genome-wide association defines more than 30 distinct susceptibility loci for Crohn's disease. Nat Genet 2008;40(8):955–962.

256. Ferguson LR, Han DY, Fraser AG, et al. Genetic factors in chronic inflammation: Single nucleotide polymorphisms in the STAT-JAK pathway, susceptibility to DNA damage and Crohn's disease in a New Zealand population. Mutat Res Mol Mech Mutagen 2010;690(1–2):108–115.

257. Zhang J-X, Song J, Wang J, Dong W-G. JAK2 rs10758669 Polymorphisms and Susceptibility to Ulcerative Colitis and Crohn's Disease: A Meta-analysis. Inflammation 2014;37(3):793–800.

258. Nahajevszky S, Andrikovics H, Batai A, et al. The prognostic impact of germline 46/1 haplotype of Janus kinase 2 in cytogenetically normal acute myeloid leukemia. Haematologica 2011;96(11):1613–1618.

259. Hermouet S, Vilaine M. The JAK2 46/1 haplotype: a marker of inappropriate myelomonocytic response to cytokine stimulation, leading to increased risk of inflammation, myeloid neoplasm, and impaired defense against infection? Haematologica 2011;96(11):1575–1579.

260. Gränsbo K, Almgren P, Sjögren M, et al. Chromosome 9p21 genetic variation explains 13% of cardiovascular disease incidence but does not improve risk prediction. J Intern Med 2013;274(3):233–240.

261. Johnson AD, Hwang S-J, Voorman A, et al. Resequencing and clinical associations of the 9p21.3 region: a comprehensive investigation in the Framingham heart study. Circulation 2013;127(7):799–810.

262. Grinfeld J, Nangalia J, Green AR. Molecular determinants of pathogenesis and clinical phenotype in myeloproliferative neoplasms. Haematologica 2017;102(1):7–17.

263. Barbui T, Thiele J, Vannucchi AM, Tefferi A. Problems and pitfalls regarding WHO-defined diagnosis of early/prefibrotic primary myelofibrosis versus essential thrombocythemia. Leukemia 2013;27(10):1953–1958.

264. Barosi G, Rosti V, Bonetti E, et al. Evidence that Prefibrotic Myelofibrosis Is Aligned along a Clinical and Biological Continuum Featuring Primary Myelofibrosis. PLoS ONE;7(4):.

265. Austin SK, Lambert JR. The JAK2V617F mutation and thrombosis. Br J Haematol 2008;143(3):307–320.

Appendices

Appendix 1: Germline single nucleotide polymorphisms genotyped

SNP	Position	Reference	Association(s)	SNP	Position	Reference	Association(s)
rs3916164	1p	[1]	MCH	rs385893	9p	[5]	PLT
rs741959	1p	[1]	MCV	rs10974900	9p	[9]	MPN
rs3811444	1q	[1,3]	RBC, PLT	rs10758658	9p	[7]	MCV
rs9660992	1q	[1,4]	MCH, MPV	rs409801	9p	[3]	PLT
rs7529925	1q	[1]	RBC	rs2236496	9p	[1,5]	MCV
rs857684	1q	[1]	MCHC	rs579459	9q	[1]	RBC
rs7551442	1q	[1]	MCHC	rs3737304	10p	[2]	MPN
rs243070	2p	[1]	MCV	rs10159477	10q	[1]	HB
rs4953318	2p	[1]	HCT	rs901683	10q	[1,8]	MCV
rs10207392	2q	[1]	MCV	rs11042125	11p	[1,8]	HB
rs4858647	3p	[2]	MPN	rs7936461	11p	[1]	HCT
rs9310736	3p	[1,5]	MCV	rs10849023	12p	[1]	MCH
rs2201862	3q	[2]	MPN	rs7312105	12p	[1]	HCT
rs11717368	3q	[1]	MCH	rs3184504	12q	[1,4]	HB, PLT
rs6776003	3q	[1]	MCV	rs11104870	12q	[1]	RBC
rs13061823	3q	[1]	MCV	rs3829290	12q	[1]	MCV
rs218238	4q	[1]	RBC	rs11627546	14q	[1]	MCV
rs13152701	4q	[1]	MCV	rs17616316	14q	[1]	MCH
rs6198	5q	[6]	MPN	rs7155454	14q	[1]	MCH
rs1408272	6p	[1,7]	MCH	rs11072566	15q	[1]	HB
rs13219787	6p	[1]	MCH	rs1532085	15q	[1]	HB
rs2097775	6p	[1]	HB	rs2572207	15q	[1]	MCV
rs6914805	6p	[1]	MCH	rs2867932	15q	[1]	MCHC
rs9369427	6p	[1]	HB	rs2271294	16q	[1]	RBC
rs9376092	6q	[2]	MPN	rs888424	17p	[1]	MCH
rs1008084	6q	[1,8]	MCH	rs8182252	17q	[1]	RBC
rs590856	6q	[1]	MCV	rs4969184	17q	[1]	HB
rs736661	6q	[1]	MCH	rs4890633	18q	[1,8]	MCH
rs12718598	7p	[1]	MCV	rs2032314	21q	[1]	HCT
rs6987853	8p	[1]	MCHC	rs5749446	22q	[1]	MCH
rs12340895	9p	[2]	MPN	rs5754217	22q	[1]	MCV

References:

1. van der Harst P, Zhang W, Mateo Leach I, et al. Seventy-five genetic loci influencing the human red blood cell. Nature 2012;492(7429):369–375.
2. Tapper W, Jones AV, Kralovics R, et al. Genetic variation at MECOM, TERT, JAK2 and HBS1L-MYB predisposes to myeloproliferative neoplasms. Nat Commun 2015;66691.
3. Gieger C, Radhakrishnan A, Cvejic A, et al. New gene functions in megakaryopoiesis and platelet formation. Nature 2011;480(7376):201–208.
4. Shameer K, Denny JC, Ding K, et al. A Genome- and Phenome-Wide Association Study to Identify Genetic Variants Influencing Platelet Count and Volume and their Pleiotropic Effects. Hum Genet;133(1)
5. Kamatani Y, Matsuda K, Okada Y, et al. Genome-wide association study of hematological and biochemical traits in a Japanese population. Nat Genet 2010;42(3):210–215.
6. Poletto V, Rosti V, Villani L, et al. A3669G polymorphism of glucocorticoid receptor is a susceptibility allele for primary myelofibrosis and contributes to phenotypic diversity and blast transformation. Blood 2012;120(15):3112–3117.
7. Ganesh SK, Zakai NA, van Rooij FJA, et al. Multiple loci influence erythrocyte phenotypes in the CHARGE Consortium. Nat Genet 2009;41(11):1191–1198.
8. Ulirsch JC, Nandakumar SK, Wang L, et al. Systematic Functional Dissection of Common Genetic Variation Affecting Red Blood Cell Traits. Cell 2016;165(6):1530–1545.
9. Hinds DA, Barnholt KE, Mesa RA, et al. Germ line variants predispose to both JAK2 V617F clonal hematopoiesis and myeloproliferative neoplasms. Blood 2016;128(8):1121–1128.

Appendix 2: R code used in statistical analysis.

```
library(CoxHD); library(survival); library(DT); library(htmlwidgets); library(IRdisplay); library(repr); library(msSurv); library(Rcpp); library(HilbertVis);
library(rms); library(data.table); library(hdp); library(ggplot2); library(lsr); library(BradleyTerry2); library(qvcalc); library(tidyr); library(survAUC)
setwd("/Users/jg23/data")

# Bradley terry modelling performed using separate file tabulating "wins/losses" for each combo of genes
btz<-read.table("btz.csv",sep=",",header=TRUE)
btm<- BTm(cbind(win1, win2), Var1, Var2, ~ Var, id = "Var", data=btz,br=TRUE,refcat="V617F")
# refcat is arbitrary
qv<-qvcalc(BTabilities(btm))
sink("TGS_output.txt",append=TRUE)
cat("Bradley Terry estimates for abilities, with quasi variances \n")
qv; sink()

rawData <- read.table("TGSZ.csv", header=T, sep=",", fill=T, na.strings=c("NA","na"))
TGSZ<-rawData
TGSZ$DeathC[which(TGSZ$AMLTC==1)]<-1
TGSZcut<-subset(TGSZ,Death>=0)

###Descriptive Stats
sink("TGS_output.txt",append=TRUE)
cat("Descriptive stats by diagnosis \n")
cat("Hb, WCC, Pl, Age by Diag \n")
aggregate(TGSZ[,c("Hb","WCC","Pl","Age")],by=list(TGSZ$Diag),FUN=median,na.rm=TRUE)
cat("Sex and Cohort by Diag \n")
table(TGSZ$Sex,TGSZ$Diag)
table(TGSZ$Cohort,TGSZ$Diag)
cat("Counts for each mutation \n")
aggregate(TGSZ[,c("JAK2", "CALR", "MPL", "JAK2e12", "TET2", "ASXL1", "DNMT3A", "PPM1D", "EZH2", "NF1", "NFE2", "SF3B1", "SRSF2",
```

```
"TP53", "U2AF1", "CBL", "MLL3", "ZRSR2", "GNAS", "KRAS", "SH2B3", "IDH2", "PTPN11", "KIT", "RB1", "BCOR", "NRAS", "CUX1", "STAG2",
"IDH1", "RUNX1", "PHF6", "GATA2", "GNB1","C1p", "C1q", "C4", "C5", "C7", "C8", "C9U", "C9g", "C11", "C12", "C13", "C14", "C17", "C18", "C19",
"C20")],by=list(TGSZ$Diag),FUN=sum,na.rm=TRUE)
cat("Linear regression for Hb \n")
summary(glm(Hb~JAK2+CALR+MPL+JAK2e12+TET2+ASXL1+DNMT3A+PPM1D+EZH2+NF1+NFE2+SF3B1+SRSF2+TP53+U2AF1+CBL+MLL3+
ZRSR2+GNAS+KRAS+SH2B3+IDH2+PTPN11+KIT+RB1+BCOR+NRAS+CUX1+STAG2+IDH1+RUNX1+PHF6+GATA2+GNB1+C1p+C1q+C4+C5+
C7+C8+C9U+C9g+C11+C12+C13+C14+C17+C18+C19+C20+Age+Sex+Diag,data=TGSZ))
cat("Linear regression for WCC \n")
summary(glm(WCC~JAK2+CALR+MPL+JAK2e12+TET2+ASXL1+DNMT3A+PPM1D+EZH2+NF1+NFE2+SF3B1+SRSF2+TP53+U2AF1+CBL+MLL
3+ZRSR2+GNAS+KRAS+SH2B3+IDH2+PTPN11+KIT+RB1+BCOR+NRAS+CUX1+STAG2+IDH1+RUNX1+PHF6+GATA2+GNB1+C1p+C1q+C4+C
5+C7+C8+C9U+C9g+C11+C12+C13+C14+C17+C18+C19+C20+Age+Sex+Diag,data=TGSZ))
cat("Linear regression for Platelet count \n")
summary(glm(Pl~JAK2+CALR+MPL+JAK2e12+TET2+ASXL1+DNMT3A+PPM1D+EZH2+NF1+NFE2+SF3B1+SRSF2+TP53+U2AF1+CBL+MLL3+
ZRSR2+GNAS+KRAS+SH2B3+IDH2+PTPN11+KIT+RB1+BCOR+NRAS+CUX1+STAG2+IDH1+RUNX1+PHF6+GATA2+GNB1+C1p+C1q+C4+C5+
C7+C8+C9U+C9g+C11+C12+C13+C14+C17+C18+C19+C20+Age+Sex+Diag,data=TGSZ))
sink()

##OR

a<-TGSZ[,c("ET","MF","PV","JAK2", "CALR", "MPL", "JAK2e12", "TET2", "ASXL1", "DNMT3A", "PPM1D", "EZH2", "NF1", "NFE2", "SF3B1",
"SRSF2", "TP53", "U2AF1", "CBL", "MLL3", "ZRSR2", "GNAS", "KRAS", "SH2B3", "IDH2", "PTPN11", "KIT", "RB1", "BCOR", "NRAS", "CUX1",
"STAG2", "IDH1", "RUNX1", "PHF6", "GATA2", "GNB1","C1p", "C1q", "C4", "C5", "C7", "C8", "C9U", "C9g", "C11", "C12", "C13", "C14", "C17",
"C18", "C19", "C20")]
a<-as.matrix(a)
 a<-t(a)%*%a; counts<-apply(a,1,max); b<-counts-a; c<-t(b); d<-nrow(TGSZ)-a-b-c
 OR<-(a*d)/(b*c)
RR<-(a/(a+b))/(c/(c+d))
SE<-sqrt(1/a+1/b+1/c+1/d)
UL<-2^(log(OR,2)+1.96*SE)
LL<- 2^(log(OR,2)-1.96*SE)
```

```
SER<-sqrt(b/(a*(a+b))+d/(c*(c+d)))
ULR<-2^(log(RR,2)+1.96*SER)
LLR<- 2^(log(RR,22)-1.96*SER)
new<-cbind(as.vector(a),as.vector(b),as.vector(c),as.vector(d))
chis<-function(dataset,a,b,c,d){
A<-0
g<-1
while(g<=nrow(dataset)){
if(is.na(dataset[g,c])){
A[g]<-NA
g<-g+1
}else{
A[g]<-fisher.test(matrix(c(dataset[g,a],dataset[g,b],dataset[g,c],dataset[g,d]), nrow=2, ncol=2))$p.value
g<-g+1}}
return(A)}
chi<-chis(new,1,2,3,4)
size<-length(a)
 chi<-matrix(chi,ncol=nrow(a),nrow=nrow(a))
 ps<-chi
 ps[lower.tri(ps,diag=TRUE)]<-NA
 ps2<-p.adjust(ps,method="BH")
ps3<-ps2
sigs<-gather(data.frame(OR),gene,OR,factor_key=TRUE)
sigs$gene2<-colnames(OR)
sigs$UL<-matrix(UL,ncol=1)
sigs$LL<-matrix(LL,ncol=1)
sigs$pvalue<-ps3
sigs<-sigs[which(ps3<0.05),]
sink("TGS_output.txt",append=TRUE)
cat("Significant associations \n")
```

```
sigs
sink()
 ps2<-matrix(ps2,ncol=nrow(a),nrow=nrow(a))
 write.table(ps2,file="Zps2.csv",sep=",")
 write.table(OR,file="ZOR.csv",sep=",")
write.table(a,file="Za.csv",sep=",")
write.table(chi,file="Zchi.csv",sep=",")
write.table(UL,file="ZUL.csv",sep=",")
write.table(LL,file="ZLL.csv",sep=",")

GWAS<-read.table("GWAS_Z.csv",sep=",",header=TRUE)
a<-GWAS[,c(7:83,91:105,108:143)]
a<-as.matrix(a)
a<-t(a)%*%a
 counts<-apply(a,1,max)
 b<-counts-a
 c<-t(b)
d<-nrow(GWAS)-a-b-c
 OR<-(a*d)/(b*c)
RR<-(a/(a+b))/(c/(c+d))
SE<-sqrt(1/a+1/b+1/c+1/d)
UL<-2^(log(OR,2)+1.96*SE)
LL<- 2^(log(OR,2)-1.96*SE)
SER<-sqrt(b/(a*(a+b))+d/(c*(c+d)))
ULR<-2^(log(RR,2)+1.96*SER)
LLR<- 2^(log(RR,22)-1.96*SER)
new<-cbind(as.vector(a),as.vector(b),as.vector(c),as.vector(d)); chis<-function(dataset,a,b,c,d){
A<-0
g<-1
while(g<=nrow(dataset)){
```

```
if(is.na(dataset[g,c])){
A[g]<-NA
g<-g+1
}else{
A[g]<-fisher.test(matrix(c(dataset[g,a],dataset[g,b],dataset[g,c],dataset[g,d]), nrow=2, ncol=2))$p.value
g<-g+1}}
return(A)}
chi<-chis(new,1,2,3,4)
size<-length(a)
 chi<-matrix(chi,ncol=nrow(a),nrow=nrow(a))
 ps<-chi
 ps[lower.tri(ps,diag=TRUE)]<-NA
 ps2<-p.adjust(ps,method="BH")
ps3<-ps2
sigs<-gather(data.frame(OR),gene,OR,factor_key=TRUE)
sigs$gene2<-colnames(OR)
sigs$UL<-matrix(UL,ncol=1)
sigs$LL<-matrix(LL,ncol=1)
sigs$pvalue<-ps3
sigs<-sigs[which(ps3<0.05),]
sink("TGS_output.txt",append=TRUE)
cat("Significant associations \n")
sigs
sink()
 ps2<-matrix(ps2,ncol=nrow(a),nrow=nrow(a))
 write.table(ps2,file="Gps2.csv",sep=",")
 write.table(OR,file="GOR.csv",sep=",")
write.table(a,file="Ga.csv",sep=",")
write.table(chi,file="Gchi.csv",sep=",")
write.table(UL,file="GUL.csv",sep=",")
```

```
write.table(LL,file="GLL.csv",sep=",")

JEP<-subset(GWAS,JAK2==1&(ET==1|PV==1))
a<-JEP[,c(7:83,91:105,108:143)]
a<-as.matrix(a)
a<-t(a)%*%a
 counts<-apply(a,1,max)
 b<-counts-a
 c<-t(b)
d<-nrow(JEP)-a-b-c
 OR<-(a*d)/(b*c)
RR<-(a/(a+b))/(c/(c+d))
SE<-sqrt(1/a+1/b+1/c+1/d)
UL<-2^(log(OR,2)+1.96*SE)
LL<- 2^(log(OR,2)-1.96*SE)
SER<-sqrt(b/(a*(a+b))+d/(c*(c+d)))
ULR<-2^(log(RR,2)+1.96*SER)
LLR<- 2^(log(RR,22)-1.96*SER)
new<-cbind(as.vector(a),as.vector(b),as.vector(c),as.vector(d))
chis<-function(dataset,a,b,c,d){
A<-0
g<-1
while(g<=nrow(dataset)){
if(is.na(dataset[g,c])){
A[g]<-NA
g<-g+1
}else{
A[g]<-fisher.test(matrix(c(dataset[g,a],dataset[g,b],dataset[g,c],dataset[g,d]), nrow=2, ncol=2))$p.value
g<-g+1}}
return(A)}
```

```
chi<-chis(new,1,2,3,4)
size<-length(a)
 chi<-matrix(chi,ncol=nrow(a),nrow=nrow(a))
 ps<-chi
 ps[lower.tri(ps,diag=TRUE)]<-NA
 ps2<-p.adjust(ps,method="BH")
ps3<-ps2
sigs<-gather(data.frame(OR),gene,OR,factor_key=TRUE)
sigs$gene2<-colnames(OR)
sigs$UL<-matrix(UL,ncol=1)
sigs$LL<-matrix(LL,ncol=1)
sigs$pvalue<-ps3
sigs<-sigs[which(ps3<0.05),]
sink("TGS_output.txt",append=TRUE)
cat("Significant associations \n")
sigs
cat("Logistic regression results \n")
summary(glm(PV~NFE2+C9+V617FHigh+rs409801+rs2867932+rs2236496+rs12340895+rs11042125+rs10974900+rs10758658+Age+Sex,family=binomi
al(link='logit'),data=JEP))
sink()

 ps2<-matrix(ps2,ncol=nrow(a),nrow=nrow(a))
 write.table(ps2,file="Jps2.csv",sep=",")
 write.table(OR,file="JOR.csv",sep=",")
write.table(a,file="Ja.csv",sep=",")
write.table(chi,file="Jchi.csv",sep=",")
write.table(UL,file="JUL.csv",sep=",")
write.table(LL,file="JLL.csv",sep=",")

# ggplot(subset(FP,gene2=="PV"),aes(x=OR,y=Order,label=pval))+geom_point()+scale_x_log10()+scale_y_discrete(limits=c("rs2236496 (C)","rs409801
```

```
(C)", "rs11042125 (T)","rs10758658 (A)","Male sex","Age >60","rs12340895 (G)","NFE2","JAK2 clone
>50%","9p"))+geom_errorbarh(aes(xmax=UL,xmin=LL,height=0.1))+geom_text(nudge_y=-0.25,nudge_x=0.2,cex=2)+theme_bw()+xlab("Odds ratio for
Polycythemia vera")+ylab("")+coord_fixed(ratio = 0.1)

#ggplot(subset(FP,gene2=="MF"),aes(x=OR,y=Order,label=pval))+geom_point()+scale_x_log10()+scale_y_discrete(limits=c("rs13219787
(A)","rs11104870 (T)","Male sex","Age >60","9p","JAK2 clone
>50%","1p","CBL","ASXL1","EZH2","SRSF2","U2AF1","7q","ZRSR2","NRAS"))+geom_errorbarh(aes(xmax=UL,xmin=LL,height=0.1))+geom_text(nu
dge_y=-0.25,nudge_x=0.2,cex=2)+theme_bw()+xlab("Odds ratio for Myelofibrosis")+ylab("")+coord_fixed(ratio = 0.1)


## Dirichlet analysis
genotypes<-TGSZ[,c("Jhet", "Jhom","CALR", "MPL", "JAK2e12", "TET2", "ASXL1", "DNMT3A", "PPM1D", "EZH2", "NF1", "NFE2", "SF3B1",
"SRSF2", "TP53", "U2AF1", "CBL", "MLL3", "ZRSR2", "GNAS", "KRAS", "SH2B3", "IDH2", "PTPN11", "KIT", "RB1", "BCOR", "NRAS", "CUX1",
"STAG2", "IDH1", "RUNX1", "PHF6", "GATA2", "GNB1","C1p", "C1q", "C4", "C5", "C7", "C8", "C11", "C12", "C13", "C14", "C17", "C18", "C19",
"C20")]
n <- ncol(genotypes)
shape <- 5
invscale <- 5
hdp <- hdp_init(ppindex=0, #index of the parent DP for initial DP
      cpindex=1, #index of alphaa and alphab for initial DP
      hh=rep(1/n,n), #params for base distn (uniform Dirichlet)
      alphaa=shape,
      alphab=invscale)
hdp <- hdp_adddp(hdp,
      numdp=nrow(genotypes), # one DP for every sample in that cancer type
      ppindex=1, # parent DP for group i is the i-th+1 overall DP because of the grandparent at position 1
      cpindex=1) # index of alphaa and alphab for each DP

# Assign the data from each patient to a child DP
hdp <- hdp_setdata(hdp = hdp, dpindex=1:nrow(genotypes)+1, data=genotypes)
```

```
# Activate the DPs with specified number of classes (signatures)
hdp <- dp_activate(hdp, 1:(nrow(genotypes)+1), initcc=5, seed=42)

burnin <- 500
postsamples <- 1000
spacebw <- 20
cpsamples <- 10
set.seed(42)
hdpChain <- hdp_posterior(hdp, #activated hdp structure
      burnin=burnin,
      n=postsamples,
      space=spacebw,
      cpiter=cpsamples)

hdpMerged <- hdp_extract_components(hdpChain, cos.merge=0.95)

posteriorSamples <- aperm(array(unlist(hdpMerged@comp_categ_counts), dim=c(dim(hdpMerged@comp_categ_counts[[1]]),
length(hdpMerged@comp_categ_counts))), c(2,3,1))
rownames(posteriorSamples) <- colnames(genotypes)
colnames(posteriorSamples) <- paste("Class", 1:ncol(posteriorSamples) -1)
posteriorMeans <- rowMeans(posteriorSamples, dim=2)
posteriorQuantiles <- apply(posteriorSamples, 1:2, quantile, c(0.025,.5,0.975), type=1)
posteriorMode <- apply(posteriorSamples, 1:2, function(x) {t <- table(x); as.numeric(names(t)[which.max(t)])})
sink("TGS_output.txt",append=TRUE)
cat("Genes by HDP group \n")
data.table(data.frame(Lesion=colnames(posteriorQuantiles), posteriorQuantiles[2,,]))
sink()

## KMs for HDP groups
```

```
pdf("HDP_KM.pdf")
par(mfrow=c(7,2), mar=c(2,2,2,1))
plot(survfit(Surv(Death/365.25,DeathC)~1,data=subset(TGSZ,MF==1&HDP=="TP53")),conf.int=FALSE,xlim=c(0,25));lines(survfit(Surv(AMLT/365.25,AMLTC)~1,data=subset(TGSZ,MF==1&HDP=="TP53")),conf.int=FALSE,xlim=c(0,25),lty=2)
plot(survfit(Surv(Death/365.25,DeathC)~1,data=subset(TGSZ,MF==0&HDP=="TP53")),conf.int=FALSE,xlim=c(0,25));lines(survfit(Surv(AMLT/365.25,AMLTC)~1,data=subset(TGSZ,MF==0&HDP=="TP53")),conf.int=FALSE,xlim=c(0,25),lty=2);lines(survfit(Surv(MFT/365.25,MFTC)~1,data=subset(TGSZ,MF==0&HDP=="TP53")),conf.int=FALSE,xlim=c(0,25),lty=3)
plot(survfit(Surv(Death/365.25,DeathC)~1,data=subset(TGSZ,MF==1&HDP=="Adv")),conf.int=FALSE,xlim=c(0,25));lines(survfit(Surv(AMLT/365.25,AMLTC)~1,data=subset(TGSZ,MF==1&HDP=="Adv")),conf.int=FALSE,xlim=c(0,25),lty=2)
plot(survfit(Surv(Death/365.25,DeathC)~1,data=subset(TGSZ,MF==0&HDP=="Adv")),conf.int=FALSE,xlim=c(0,25));lines(survfit(Surv(AMLT/365.25,AMLTC)~1,data=subset(TGSZ,MF==0&HDP=="Adv")),conf.int=FALSE,xlim=c(0,25),lty=2);lines(survfit(Surv(MFT/365.25,MFTC)~1,data=subset(TGSZ,MF==0&HDP=="Adv")),conf.int=FALSE,xlim=c(0,25),lty=3)
plot(survfit(Surv(Death/365.25,DeathC)~1,data=subset(TGSZ,MF==1&HDP=="CALR")),conf.int=FALSE,xlim=c(0,25));lines(survfit(Surv(AMLT/365.25,AMLTC)~1,data=subset(TGSZ,MF==1&HDP=="CALR")),conf.int=FALSE,xlim=c(0,25),lty=2)
plot(survfit(Surv(Death/365.25,DeathC)~1,data=subset(TGSZ,MF==0&HDP=="CALR")),conf.int=FALSE,xlim=c(0,25));lines(survfit(Surv(AMLT/365.25,AMLTC)~1,data=subset(TGSZ,MF==0&HDP=="CALR")),conf.int=FALSE,xlim=c(0,25),lty=2);lines(survfit(Surv(MFT/365.25,MFTC)~1,data=subset(TGSZ,MF==0&HDP=="CALR")),conf.int=FALSE,xlim=c(0,25),lty=3)
plot(survfit(Surv(Death/365.25,DeathC)~1,data=subset(TGSZ,MF==1&HDP=="MPL")),conf.int=FALSE,xlim=c(0,25));lines(survfit(Surv(AMLT/365.25,AMLTC)~1,data=subset(TGSZ,MF==1&HDP=="MPL")),conf.int=FALSE,xlim=c(0,25),lty=2)
plot(survfit(Surv(Death/365.25,DeathC)~1,data=subset(TGSZ,MF==0&HDP=="MPL")),conf.int=FALSE,xlim=c(0,25));lines(survfit(Surv(AMLT/365.25,AMLTC)~1,data=subset(TGSZ,MF==0&HDP=="MPL")),conf.int=FALSE,xlim=c(0,25),lty=2);lines(survfit(Surv(MFT/365.25,MFTC)~1,data=subset(TGSZ,MF==0&HDP=="MPL")),conf.int=FALSE,xlim=c(0,25),lty=3)
plot(survfit(Surv(Death/365.25,DeathC)~1,data=subset(TGSZ,MF==1&HDP=="Jhom")),conf.int=FALSE,xlim=c(0,25));lines(survfit(Surv(AMLT/365.25,AMLTC)~1,data=subset(TGSZ,MF==1&HDP=="Jhom")),conf.int=FALSE,xlim=c(0,25),lty=2)
plot(survfit(Surv(Death/365.25,DeathC)~1,data=subset(TGSZ,MF==0&HDP=="Jhom")),conf.int=FALSE,xlim=c(0,25));lines(survfit(Surv(AMLT/365.25,AMLTC)~1,data=subset(TGSZ,MF==0&HDP=="Jhom")),conf.int=FALSE,xlim=c(0,25),lty=2);lines(survfit(Surv(MFT/365.25,MFTC)~1,data=subset(TGSZ,MF==0&HDP=="Jhom")),conf.int=FALSE,xlim=c(0,25),lty=3)
plot(survfit(Surv(Death/365.25,DeathC)~1,data=subset(TGSZ,MF==1&HDP=="Jhet")),conf.int=FALSE,xlim=c(0,25));lines(survfit(Surv(AMLT/365.25,AMLTC)~1,data=subset(TGSZ,MF==1&HDP=="Jhet")),conf.int=FALSE,xlim=c(0,25),lty=2)
plot(survfit(Surv(Death/365.25,DeathC)~1,data=subset(TGSZ,MF==0&HDP=="Jhet")),conf.int=FALSE,xlim=c(0,25));lines(survfit(Surv(AMLT/365.25,A
```

```
MLTC)~1,data=subset(TGSZ,MF==0&HDP=="Jhet")),conf.int=FALSE,xlim=c(0,25),lty=2);lines(survfit(Surv(MFT/365.25,MFTC)~1,data=subset(TGSZ,
MF==0&HDP=="Jhet")),conf.int=FALSE,xlim=c(0,25),lty=3)
plot(survfit(Surv(Death/365.25,DeathC)~1,data=subset(TGSZ,MF==1&HDP=="Nil")),conf.int=FALSE,xlim=c(0,25));lines(survfit(Surv(AMLT/365.25,AM
LTC)~1,data=subset(TGSZ,MF==1&HDP=="Nil")),conf.int=FALSE,xlim=c(0,25),lty=2)
plot(survfit(Surv(Death/365.25,DeathC)~1,data=subset(TGSZ,MF==0&HDP=="Nil")),conf.int=FALSE,xlim=c(0,25));lines(survfit(Surv(AMLT/365.25,AM
LTC)~1,data=subset(TGSZ,MF==0&HDP=="Nil")),conf.int=FALSE,xlim=c(0,25),lty=2);lines(survfit(Surv(MFT/365.25,MFTC)~1,data=subset(TGSZ,MF
==0&HDP=="Nil")),conf.int=FALSE,xlim=c(0,25),lty=3)
dev.off()

##p-values, median survival and 10-yr outcomes for HDP groups
TGSZ$HDP<-relevel(TGSZ$HDP,ref="Jhet")
sink("TGS_output.txt",append=TRUE)
coxph(Surv(Death,DeathC)~HDP,subset(TGSZ,MF==0))
coxph(Surv(Death,DeathC)~HDP,subset(TGSZ,MF==1))
coxph(Surv(MFT,MFTC)~HDP,subset(TGSZ,MF==0))
coxph(Surv(EFS,EFSC)~HDP,subset(TGSZ,MF==0))
coxph(Surv(AMLT,AMLTC)~HDP,subset(TGSZ,MF==0))
coxph(Surv(AMLT,AMLTC)~HDP,subset(TGSZ,MF==1))

survfit(Surv(Death,DeathC)~HDP,subset(TGSZ,MF==0))
survfit(Surv(Death,DeathC)~HDP,subset(TGSZ,MF==1))
survfit(Surv(MFT,MFTC)~HDP,subset(TGSZ,MF==0))
survfit(Surv(EFS,EFSC)~HDP,subset(TGSZ,MF==0))
survfit(Surv(AMLT,AMLTC)~HDP,subset(TGSZ,MF==0))
survfit(Surv(AMLT,AMLTC)~HDP,subset(TGSZ,MF==1))

data.frame(levels(TGSZ$HDP),summary(survfit(Surv(Death,DeathC)~HDP,subset(TGSZ,MF==0)),times=c(10*365.25))$surv)
data.frame(levels(TGSZ$HDP),summary(survfit(Surv(Death,DeathC)~HDP,subset(TGSZ,MF==1)),times=c(10*365.25))$surv)
data.frame(levels(TGSZ$HDP),summary(survfit(Surv(MFT,MFTC)~HDP,subset(TGSZ,MF==0)),times=c(10*365.25))$surv)
data.frame(levels(TGSZ$HDP),summary(survfit(Surv(EFS,EFSC)~HDP,subset(TGSZ,MF==0)),times=c(10*365.25))$surv)
```

```
data.frame(levels(TGSZ$HDP),summary(survfit(Surv(AMLT,AMLTC)~HDP,subset(TGSZ,MF==0)),times=c(10*365.25))$surv)
data.frame(levels(TGSZ$HDP),summary(survfit(Surv(AMLT,AMLTC)~HDP,subset(TGSZ,MF==1)),times=c(10*365.25))$surv)
sink()

###Descriptive Stats by HDP
sink("TGS_output.txt",append=TRUE)
cat("Descriptive stats by HDP \n")
cat("Hb, WCC, Pl, Age by HDP \n")
aggregate(TGSZ[,c("Hb","WCC","Pl","Age")],by=list(TGSZ$HDP),FUN=median,na.rm=TRUE)
cat("Sex and Cohort by HDP \n")
table(TGSZ$Sex,TGSZ$HDP)
table(TGSZ$Cohort,TGSZ$HDP)
table(TGSZ$Diag,TGSZ$HDP)
cat("Counts for each mutation \n")
aggregate(TGSZ[,c("JAK2", "CALR", "MPL", "JAK2e12", "TET2", "ASXL1", "DNMT3A", "PPM1D", "EZH2", "NF1", "NFE2", "SF3B1", "SRSF2",
"TP53", "U2AF1", "CBL", "MLL3", "ZRSR2", "GNAS", "KRAS", "SH2B3", "IDH2", "PTPN11", "KIT", "RB1", "BCOR", "NRAS", "CUX1", "STAG2",
"IDH1", "RUNX1", "PHF6", "GATA2", "GNB1","C1p", "C1q", "C4", "C5", "C7", "C8", "C9U", "C9g", "C11", "C12", "C13", "C14", "C17", "C18", "C19",
"C20")],by=list(TGSZ$HDP),FUN=sum,na.rm=TRUE)
summary(glm(Hb~HDP+Age+Sex+Diag,data=TGSZ))
summary(glm(Hb~HDP+Age+Sex,data=subset(TGSZ,ET==1)))
summary(glm(Hb~HDP+Age+Sex,data=subset(TGSZ,PV==1)))
summary(glm(Hb~HDP+Age+Sex,data=subset(TGSZ,MF==1)))
summary(glm(WCC~HDP+Age+Sex+Diag,data=TGSZ))
summary(glm(WCC~HDP+Age+Sex,data=subset(TGSZ,ET==1)))
summary(glm(WCC~HDP+Age+Sex,data=subset(TGSZ,PV==1)))
summary(glm(WCC~HDP+Age+Sex,data=subset(TGSZ,MF==1)))
summary(glm(Pl~HDP+Age+Sex+Diag,data=TGSZ))
summary(glm(Pl~HDP+Age+Sex,data=subset(TGSZ,ET==1)))
summary(glm(Pl~HDP+Age+Sex,data=subset(TGSZ,PV==1)))
summary(glm(Pl~HDP+Age+Sex,data=subset(TGSZ,MF==1)))
```

```
sink()

## Data prep for individual models
rawData$MFT[rawData$MFTC==0 | is.na(rawData$MFTC)] <- NA
rawData$AMLT[rawData$AMLTC==0 | is.na(rawData$AMLTC)] <- NA
rawData$CPT <- rawData$Lcen
# This takes into account patients that start in MF
rawData$MFT[rawData$MF==1] <- rawData$Lcen[rawData$MF==1]
rawData$CPT[rawData$MF==1] <- NA

rawData$Diagnosis <- rep(0, nrow(rawData))
# Change gender to an integer
rownames(rawData) <- rawData$UPN
Cohorts <- MakeInteger(rawData$Cohort)
rawData[colnames(Cohorts)] <- Cohorts
rawData$Cohort <- NULL
rawData$Sex <- as.integer(rawData$Sex)
splitPatients <- function(progsTimes, Z){
	Zdup <- lapply(1:nrow(progsTimes), function(i){
		id <- progsTimes[i,]$id
		patient_data <- unlist(Z[rownames(Z) == id,])
		return(patient_data)
	})
	return(as.data.frame(do.call("rbind",Zdup)))
}
progressionTimes <- function(data, startName, endName, transNames, outcomeName){
	progress_data <- NULL
	for (i in c(1:nrow(data))){
		t <- c(as.numeric(data[i,c(transNames,endName)]) - as.numeric(data[i,startName]))
		order_idx <- order(t, na.last=NA)
```

```
            ordered_t <- c(0,t[order_idx])
            start_t <- ordered_t[-length(ordered_t)]
            end_t <- ordered_t[-1]
            trans_group <- matrix(0, length(order_idx), length(c(startName, transNames)))
            for (r in 1:length(order_idx))
                  trans_group[r, c(1,order_idx+1)[r]] <- 1
            colnames(trans_group) <- c(startName, transNames)
            outcome <- rep(0, length(start_t))
            outcome[length(outcome)] <- data[i,outcomeName]
            s <- data.frame(id = rownames(data)[i], trans_group, start.time = start_t, end.time = end_t, outcome=outcome)
            progress_data <- rbind(progress_data, s)
      }
      return(progress_data)
}
dataGroups <- list(
   Genetics = c("JAK2", "CALR1","CALR2", "MPL", "JAK2e12", "TET2", "ASXL1", "DNMT3A", "PPM1D", "EZH2", "NF1", "NFE2", "SF3B1",
"SRSF2", "TP53", "U2AF1", "CBL", "MLL3", "ZRSR2", "GNAS", "KRAS", "SH2B3", "IDH2", "PTPN11", "KIT", "RB1", "BCOR", "NRAS", "CUX1",
"STAG2", "IDH1", "RUNX1", "PHF6", "GATA2", "GNB1"),
   Cytogenetics = c("C1p", "C1q", "C4", "C5", "C7", "C8", "C9U", "C9g", "C11", "C14", "C17", "C18", "C19", "C20"),
   Demographics = c("Age", "Sex"),
   Clinical = c("Splen", "Hb", "WCC", "Pl", "PV", "ET","MF","PriorThrom"),
Nuisance = c(colnames(Cohorts))
)

mustHaveCols <- c("Death","DeathC")
for (col in mustHaveCols)
   rawData <- rawData[!is.na(rawData[[col]]),]

aml_corrections <- !is.na(rawData$AMLTC) & rawData$AMLTC & (rawData$Death - rawData$AMLT) < 1 & !is.na(rawData$AMLT)
rawData$Death[aml_corrections] <- rawData$AMLT[aml_corrections] + 1
```

```
mf_corrections <- !is.na(rawData$MFTC) & rawData$MFTC & (rawData$Death - rawData$MFT) < 1 & !is.na(rawData$MFT)
rawData$Death[mf_corrections] <- rawData$MFT[mf_corrections] + 1
time_mf_cp <- rawData$MFT[rawData$MF == 0]/365
time_aml_cp <- rawData$AMLT[rawData$MF == 0]/365
time_death_cp <- pmin(rawData$Death[rawData$MF == 0]/365, time_aml_cp, na.rm = T)
outcome_from_cp <- as.integer(rawData$DeathC[rawData$MF == 0] | rawData$AMLTC[rawData$MF == 0])
time_mf_mf <- rawData$MFT[rawData$MF != 0]/365
time_aml_mf <- rawData$AMLT[rawData$MF != 0]/365
time_death_mf <- pmin(rawData$Death[rawData$MF != 0]/365, time_aml_mf, na.rm = T)
outcome_from_mf <- as.integer(rawData$DeathC[rawData$MF != 0] | rawData$AMLTC[rawData$MF != 0])
timeStage <- progressionTimes(rawData, "Diagnosis", "Death", c("CPT", "MFT", "AMLT"), "DeathC")
rawZ <- rawData[unlist(dataGroups)]
ZStdize <- StandardizeMagnitude(rawZ)
colnames(ZStdize) <- colnames(rawZ)
poorMansImpute <- function(x) {x[is.na(x)] <- mean(x, na.rm=TRUE); return(x)}
ZImpute <- as.data.frame(sapply(ZStdize, poorMansImpute))
rownames(ZImpute) <- rownames(ZStdize)
Zdup <- splitPatients(timeStage, ZImpute)
all_data <-list(id=timeStage$id,
       start.time=timeStage$start.time/365,
       end.time=timeStage$end.time/365,
       outcome=timeStage$outcome,
       progression=timeStage[c("Diagnosis", c("CPT", "MFT", "AMLT"))],
       Z=Zdup)
which.mu <- c()
new_split_patient <- which(all_data$progression$Diagnosis == 1)
split_patient_diagnosis <- c()
cp_or_mf <- "CP"
for (i in (1:nrow(all_data$progression))) {
   if (i %in% new_split_patient) {
```

```
        if (as.integer(all_data$progression[i+1,]$CPT) == 1){
            cp_or_mf <- "CP"
        }
        if (as.integer(all_data$progression[i+1,]$MFT) == 1){
            cp_or_mf <- "MF"
        }
    }
    split_patient_diagnosis <- c(split_patient_diagnosis, cp_or_mf)
}
wpresampling <- which(all_data$progression$Diagnosis == 1)
all_data$sampling.time <- all_data$start.time[wpresampling+1] - all_data$start.time[wpresampling]
# Remove following line if you want the start time to go from sampling
all_data$start.time[wpresampling+1] <- all_data$start.time[wpresampling]
###
all_data$Z <- all_data$Z[-wpresampling,]
all_data$id <- all_data$id[-wpresampling]
all_data$progression <- all_data$progression[-wpresampling,]
all_data$start.time <- all_data$start.time[-wpresampling]
all_data$end.time <- all_data$end.time[-wpresampling]
all_data$outcome <- all_data$outcome[-wpresampling]
all_data$split_patient_diagnosis <- all_data$split_patient_diagnosis[-wpresampling]
split_patient_diagnosis <- split_patient_diagnosis[-wpresampling]
mf_data <-list(id=all_data$id[split_patient_diagnosis=="MF"],
    start.time=all_data$start.time[split_patient_diagnosis=="MF"],
    end.time=all_data$end.time[split_patient_diagnosis=="MF"],
    outcome=all_data$outcome[split_patient_diagnosis=="MF"],
    progression=all_data$progression[split_patient_diagnosis=="MF",],
    Z=all_data$Z[split_patient_diagnosis=="MF",])
cp_data <-list(id=all_data$id[split_patient_diagnosis=="CP"],
    start.time=all_data$start.time[split_patient_diagnosis=="CP"],
```

```
    end.time=all_data$end.time[split_patient_diagnosis=="CP"],
    outcome=all_data$outcome[split_patient_diagnosis=="CP"],
    progression=all_data$progression[split_patient_diagnosis=="CP",],
    Z=all_data$Z[split_patient_diagnosis=="CP",])
dataList <- list(
    Genetics = all_data$Z[,dataGroups$Genetics],
    Cytogenetics = all_data$Z[,dataGroups$Cytogenetics],
    Demographics = all_data$Z[,dataGroups$Demographics],
    Clinical = all_data$Z[,dataGroups$Clinical],
    Nuisance = all_data$Z[,dataGroups$Nuisance]
)
groups <- unlist(sapply(names(dataList), function(x) rep(x, ncol(dataList[[x]]))))
groups <- factor(groups)
names(groups) <- colnames(all_data$Z)
table(groups)
wcp <- which(cp_data$progression$CPT == 1) # Sub patients that are in CP
wcpaml <- intersect(wcp+1, which(cp_data$progression$AMLT == 1)) # Sub patients that have progressed to AML from CP
wcpmf <- which(cp_data$progression$MFT == 1) # Sub patients that have progressed to MF from CP
wcpmfaml <- intersect(wcpmf+1, which(cp_data$progression$AMLT == 1)) # Sub patients that have progressed to AML from Secondary MF
length(wcp)
length(wcpaml)
length(wcpmf)
length(wcpmfaml)
wmf <- which(mf_data$progression$MFT == 1) # Sub patients that are in MF
wmfaml <- intersect(wmf+1, which(mf_data$progression$AMLT == 1)) # Sub patients that have progressed to AML from primary MF
length(wmf)
length(wmfaml)
test_os <- Surv(cp_data$start.time[wcp], time_death_cp, outcome_from_cp)
cp_or_mf <- union(wcp, wcpmf)
o <- order(cp_or_mf)
```

```
cp_or_mf <- cp_or_mf[o]
os_from_cp <- Surv(cp_data$start.time[cp_or_mf], cp_data$end.time[cp_or_mf], as.integer(cp_data$outcome[cp_or_mf] |
cp_data$progression$AMLT[cp_or_mf+1]==1))
#os_from_cp <- Surv(cp_data$end.time[cp_or_mf] - cp_data$start.time[cp_or_mf], as.integer(cp_data$outcome[cp_or_mf] |
cp_data$progression$AMLT[cp_or_mf+1]==1))

plot(survfit(os_from_cp~1), xlim=c(0, 25), xlab="Yrs")
lines(survfit(test_os~1), col="red")
surv_cp <- Surv(cp_data$start.time, cp_data$end.time, cp_data$outcome)[wcp,]
#surv_cp <- Surv(cp_data$end.time - cp_data$start.time, cp_data$outcome)[wcp,]

params_used_cp <- colSums(cp_data$Z[wcp,])!=0
cp_fit <- CoxRFX(cp_data$Z[wcp, params_used_cp], surv_cp, groups=groups[params_used_cp], which.mu=which.mu, nu=1, sigma0=0.01, max.iter=200)
o <- rep(0,length(cp_data$outcome))
o[intersect(wcp,wcpaml-1)] <- 1
cp_to_aml_surv <- Surv(cp_data$start.time, cp_data$end.time, o)[wcp,]
#cp_to_aml_surv <- Surv(cp_data$end.time - cp_data$start.time, o)[wcp,]
params_used_cp_to_aml <- colSums(cp_data$Z[wcp,])!=0
cp_to_aml_fit <- CoxRFX(cp_data$Z[wcp, params_used_cp_to_aml], cp_to_aml_surv, groups=groups[params_used_cp_to_aml], which.mu=which.mu,
nu=1, sigma0=0.01, max.iter=200)
o <- rep(0,length(cp_data$outcome))
o[intersect(wcp,wcpmf-1)] <- 1
cp_to_mf_surv <- Surv(cp_data$start.time, cp_data$end.time, o)[wcp,]
#cp_to_mf_surv <- Surv(cp_data$end.time - cp_data$start.time, o)[wcp,]
params_used_cp_to_mf <- colSums(cp_data$Z[wcp,])!=0
cp_to_mf_fit <- CoxRFX(cp_data$Z[wcp, params_used_cp_to_mf], cp_to_mf_surv, groups=groups[params_used_cp_to_mf], which.mu=which.mu, nu=1,
sigma0=0.01, max.iter=200)
combined_Z <- rbind(cp_data$Z[wcpmf,], mf_data$Z[wmf,])
combined_Z$AML_from_MF_Secondary <- c(rep(0, length(wcpmf)), rep(1, length(wmf)))
#dataGroups$Nuisance <- c(dataGroups$Nuisance, "AML_from_MF_Secondary")
```

```
#dataList$Nuisance <- combined_Z[,dataGroups$Nuisance]
dataGroups$Clinical <- c(dataGroups$Clinical, "AML_from_MF_Secondary")
dataList$Clinical <- combined_Z[,dataGroups$Clinical]
combined_groups <- unlist(sapply(names(dataList), function(x) rep(x, ncol(dataList[[x]]))))
combined_groups <- factor(combined_groups)
names(combined_groups) <- colnames(combined_Z)
rebase_start_time <- c(rep(0, length(cp_data$start.time[wcpmf])), mf_data$start.time[wmf])
rebase_end_time <- c((cp_data$end.time[wcpmf] - cp_data$start.time[wcpmf]), mf_data$end.time[wmf])
surv_mf <- Surv(rebase_start_time, rebase_end_time, c(cp_data$outcome[wcpmf], mf_data$outcome[wmf]))
#surv_mf <- Surv(rebase_end_time - rebase_start_time, c(cp_data$outcome[wcpmf], mf_data$outcome[wmf]))
params_used_mf <- colSums(combined_Z)!=0
combined_groups[params_used_mf]
mf_fit <- CoxRFX(combined_Z[params_used_mf], surv_mf, groups=combined_groups[params_used_mf], which.mu=which.mu, nu=1, sigma0=0.01,
max.iter=200)
rebase_start_time <- c(rep(0, length(cp_data$start.time[wcpmf])), mf_data$start.time[wmf])
rebase_end_time <- c(cp_data$end.time[wcpmf] - cp_data$start.time[wcpmf], mf_data$end.time[wmf])
o_cp <- rep(0,length(cp_data$outcome))
o_cp[intersect(wcpmf,wcpmfaml-1)] <- 1
o_mf <- rep(0,length(mf_data$outcome))
o_mf[intersect(wmf,wmfaml-1)] <- 1
o <- c(o_cp[wcpmf], o_mf[wmf])
mf_to_aml_surv <- Surv(rebase_start_time, rebase_end_time, o)
#mf_to_aml_surv <- Surv(rebase_end_time - rebase_start_time, o)
params_used_mf_to_aml <- colSums(combined_Z)!=0
mf_to_aml_fit <- CoxRFX(combined_Z[params_used_mf_to_aml], mf_to_aml_surv, groups=combined_groups[params_used_mf_to_aml],
which.mu=which.mu, nu=1, sigma0=0.01, max.iter=200)
o <- rep(0,length(cp_data$outcome)+length(mf_data$outcome))
o[intersect(wcp,wcpaml-1)] <- 1
o[length(cp_data$outcome)+intersect(wmf,wmfaml-1)] <- 1
o[intersect(wcp,wcpmfaml-2)] <- 1
```

```
cpormf<-c(wcp,length(cp_data$outcome)+wmf)
aml_start_time<-c(cp_data$start.time,mf_data$start.time)
aml_end_time<-c(cp_data$end.time,mf_data$end.time)
aml_end_time[intersect(wcp,wcpmfaml-2)]<-aml_end_time[wcpmfaml]
combinedaml_surv <- Surv(aml_start_time, aml_end_time, o)[cpormf,]
aml_fit <- CoxRFX(rbind(cp_data$Z,mf_data$Z)[cpormf,], combinedaml_surv, groups=groups, which.mu=which.mu, nu=1, sigma0=0.01, max.iter=200)
d <- sapply(1:length(wcp), function(i){
        i <<- i
        t <- c(time_mf_cp[i], time_death_cp[i]) - cp_data$start.time[wcp][i]
        o <- order(t, na.last=NA)
        stages <- c(1:2,0)
        r <- stages[c(1, o+1)]
        if(outcome_from_cp[i] & is.na(time_aml_cp[i]))
          r[length(r)] <- r[length(r)-1] +2
        if(!is.na(time_aml_cp[i]))
          r[length(r)] <- r[length(r)-1] +4
        tt <- c(cp_data$start.time[wcp][i],t[o]+cp_data$start.time[wcp][i])
        if(length(o)==0)
          return(c(rep(NA,7),i))
        s <- cbind(id=i, start=tt[-length(tt)],  stop=tt[-1], start.stage=r[-length(r)], end.stage=r[-1])[diff(tt)!=0,]
        # s <- cbind(time1 = tt[-length(tt)], time2=tt[-1], death=c(rep(0, length(o)-1), clinicalData$Status[i]), outer(0:(length(o)-1), r[-3], `>=`)+0,
i=i)[diff(tt)!=0,]
        return(s)
      })
d <- as.data.frame(do.call("rbind",d))
nodes <- as.character(1:6)
edges <- list(`1`=list(edges=c("2","3","5")), `2`=list(edges=c("4","6")), `3`=list(edges=NULL), `4`=list(edges=NULL),
`5`=list(edges=NULL),`6`=list(edges=NULL))
struct <-  new("graphNEL", nodes = nodes, edgeL = edges, edgemode = "directed")
msurv <- msSurv(d, struct, bs = FALSE, LT=TRUE)
```

```
y <- t(apply(cbind(1,-msurv@ps[,c(3:6, 2:1)]),1,cumsum))
par(mar=c(3,3,1,1), bty="n", mgp=c(2,.5,0), las=1)
plot(msurv@et, y[,1], ylim=c(0,1), type="s",lty=0, xlab="Time after diagnosis (years)", ylab="Fraction of patients", xlim=c(0,25), xaxs="i", yaxs="i")
steps <- function(x, type="s") rep(x, each=2)[if(type=="s") -1 else -2*length(x)]
x <- steps(msurv@et, type="S")
for(i in 1:6)
   polygon(c(x, rev(x)), c(steps(y[,i]), rev(steps(y[,i+1])) ), col=c(brewer.pal(5,"Pastel1")[c(1,2,3,4,5)],"#DDDDDD")[i], border=NA)
abline(h=seq(0,1,.2), col='white', lty=3)
abline(v=seq(0,25,1), col='white', lty=3)
#plot(survfit(os_from_cp ~ 1), xlim=c(0, 25), ylim=c(00, 1))
lines(x, steps(y[,5]), lwd=2, col="black")
#lines(survfit(test2_surv ~ 1), col="green")
w <- which.min(abs(msurv@et-25))
text(x=par("usr")[2], y= y[w,-7]+diff(y[w,])/2, labels=c(c("Death in post MF AML","Death in post CP AML","Death in MF","Death in CP")[4:1],"Alive in
MF","Alive in CP"), pos=2)
length(cp_or_mf)
MultiRFX5 <- function(coxRFXcpd, coxRFXamld, coxRFXmf, coxRFXpmfd, coxRFXpmfamld, data, x =365, tdMfBaseline = rep(1, ceiling(max(x))+1),
tdMfAmlBaseline = rep(1, ceiling(max(x))+1), tdPrmBaseline = rep(1, ceiling(max(x))+1), tdOsBaseline = rep(1, ceiling(max(x))+1), ciType="analytical",
stage0="CP"){
   cppFunction('NumericVector computeHierarchicalSurvival(NumericVector x, NumericVector diffS0, NumericVector S1Static, NumericVector
haz1TimeDep) {
                int xLen = x.size();
                double h;
                NumericVector overallSurvival(xLen);
                for(int i = 0; i < xLen; ++i) overallSurvival[i] = 1;
                for(int j = 1; j < xLen; ++j) if(diffS0[j-1] != 0){
                h = haz1TimeDep[j-1];
                for(int i = j; i < xLen; ++i){
                overallSurvival[i] += diffS0[j-1] * (1-pow(S1Static[i-j], h));
                }
```

```
        }
        return overallSurvival;
        }')
## Step 1: Compute KM survival curves and log hazard
getS <- function(coxRFX, data, max.x=5000) {
   if(!is.null(coxRFX$na.action)) coxRFX$Z <- coxRFX$Z[-coxRFX$na.action,]
   data <- as.matrix(data[,match(colnames(coxRFX$Z),colnames(data)), drop=FALSE])
   r <- PredictRiskMissing(coxRFX, data, var="var2")
   H0 <- basehaz(coxRFX, centered = FALSE)
   hazardDist <- splinefun(H0$time, H0$hazard, method="monoH.FC")
   x <- c(0:ceiling(max.x/25))*25/365
   S <- exp(-hazardDist(x))
   return(list(S=S, r=r, x=x, hazardDist=hazardDist, r0 = coxRFX$means %*% coef(coxRFX)))
}
kmMf <- getS(coxRFX = coxRFXmf, data = data, max.x=max(x))
kmCpd <- getS(coxRFX = coxRFXcpd, data = data, max.x=max(x))
kmAmld <- getS(coxRFX = coxRFXamld, data = data, max.x=max(x))
data$AML_from_MF_Secondary <- rep(ifelse(stage0=="CP", 1, 0), nrow(data))
kmPmfd <- getS(coxRFX = coxRFXpmfd, data = data, max.x=max(x))
kmPmfamld <- getS(coxRFX = coxRFXpmfamld, data = data, max.x=max(x))
xx <- 0:ceiling(max(x)/25)
sapply(1:nrow(data), function(i){
        # Adjust curves for competing risks, accounting for hazard
        # CP Death <-------- CP --------> AML Death
        #                 |
        #                 |
        #                 |
        #                 |
        #                 V
        # MF Death <-------- MF --------> AML Death
```

```
        mfpAbs <-  cumsum(c(1,diff(kmMf$S^exp(kmMf$r[i,1]))) * (kmCpd$S ^ exp(kmCpd$r[i,1])) * (kmAmld$S ^ exp(kmAmld$r[i,1])))
        cpdAbs  <- cumsum(c(1,diff(kmCpd$S^exp(kmCpd$r[i,1]))) * (kmAmld$S ^ exp(kmAmld$r[i,1])) * (kmMf$S ^ exp(kmMf$r[i,1]))) ## array
times x nrow(data)
        amldAbs  <- cumsum(c(1,diff(kmAmld$S^exp(kmAmld$r[i,1]))) * (kmMf$S ^ exp(kmMf$r[i,1])) * (kmCpd$S ^ exp(kmCpd$r[i,1])))
        mfdAbs <- cumsum(c(1,diff(kmPmfd$S^exp(kmPmfd$r[i,1]))) * (kmPmfamld$S ^ exp(kmPmfamld$r[i,1])))
        mfamldAbs <- cumsum(c(1,diff(kmPmfamld$S^exp(kmPmfamld$r[i,1]))) * (kmPmfd$S ^ exp(kmPmfd$r[i,1]))) ## array times x nrow(data)
        ### Overall survival from cronic phase
        mfamldCp <- computeHierarchicalSurvival(x = xx, diffS0 = diff(mfpAbs), S1Static = mfamldAbs, haz1TimeDep = tdMfAmlBaseline)
        mfdCp <- computeHierarchicalSurvival(x = xx, diffS0 = diff(mfpAbs), S1Static = mfdAbs, haz1TimeDep = tdMfBaseline)
        ### Overall survival from starting in MF
        mfamldMf <- mfamldAbs
        mfdMf <- mfdAbs
        cbind(
          deathInCpFromCp=1 - cpdAbs,
          deathInMfFromCp=1 - mfdCp,
          deathInAmlFromCp=1 - amldAbs,
          deathInMfamlFromCp=1 - mfamldCp,
          aliveInMfFromCp=1 - mfpAbs - (1 - mfdCp) - (1 - mfamldCp),
          deathInMfFromMF=1 - mfdMf,
          deathInAmlFromMF=1 - mfamldMf
        )
      }, simplify='array')
}

save(all_data, cp_data, mf_data, groups, MultiRFX5, dataGroups, dataList, file="loo.RData")
smf <- mf_fit$surv
coxphMf <- coxph(smf[1:length(wcpmf)] ~ pspline(time0, df=3) + I(offset), data=data.frame(time0=cp_data$start.time[wcpmf] - cp_data$start.time[wcpmf
- 1], offset=mf_fit$linear.predictors[1:length(wcpmf)]))
tdMfBaseline <- pmin(50,pmax(0.1,exp(predict(coxphMf, newdata=data.frame(time0=seq(0,25,25/365),
offset=mean(mf_fit$linear.predictors[1:length(wcpmf)])))))
```

```
plot(seq(0,25,25/365),tdMfBaseline, log="y")

# saml <- mf_to_aml_fit$surv
# coxphMfAml <- coxph(saml[1:length(wcpmf)] ~ pspline(time0, df=3) + I(offset), data=data.frame(time0=cp_data$start.time[wcpmf] -
cp_data$start.time[wcpmf - 1], offset=mf_to_aml_fit$linear.predictors[1:length(wcpmf)]))
# tdMfAmlBaseline <-  pmin(10,pmax(0.1,exp(predict(coxphMfAml, newdata=data.frame(time0=seq(0,25,25/365),
offset=mean(mf_to_aml_fit$linear.predictors[1:length(wcpmf)])))))
# plot(seq(0,25,25/365),tdMfAmlBaseline, log="y")
fiveStagePredictedCP <- MultiRFX5(cp_fit, aml_fit, cp_to_mf_fit, mf_fit, aml_fit, cp_data$Z[wcp,], x=365*25)#, tdMfBaseline=tdMfBaseline,
tdMfAmlBaseline=tdMfAmlBaseline)
fiveStagePredictedMF <- MultiRFX5(cp_fit, cp_to_aml_fit, cp_to_mf_fit, mf_fit, mf_to_aml_fit, mf_data$Z[wmf,], x=365*25)#,
tdMfBaseline=tdMfBaseline, tdMfAmlBaseline=tdMfAmlBaseline)

sedimentPlot <- function(Y, x=1:nrow(Y)*25/365, y0=0, y1=NULL, col=1:ncol(Y), ...){
   Z <- cbind(t(apply(cbind(y0,Y),1,cumsum)),y1)
   plot(x,Z[,1], xlim=range(x), ylim=range(Z), lty=0, pch=NA,...)
   for(i in 2:ncol(Z))
      polygon(c(x,rev(x)), c(Z[,i-1],rev(Z[,i])), border=NA, col=col[i-1])
}
lineStageCP <- function(mf_date, Date_LF, ERDate, aml_date, Status, y=0, col=1:5, pch.trans=19, pch.end=19, ...){
   xpd <- par("xpd")
   par(xpd=NA)
   t <- as.numeric(c(mf_date, Date_LF) - ERDate )
   w <- !is.na(t)
   o <- order(t)
   to <- pmin(t[o], par("usr")[2])
   l <- length(to)
   segments(c(0,to[-l]), rep(y,l), to, rep(y,l), col=col, lend=1, ...)
   status <- 1
   if(!is.na(mf_date))
```

```
      status <- status + 1
    if(Status==1 & is.na(aml_date))
      status <- status + 2
    if(!is.na(aml_date))
      status <- status + 4
    x <- ifelse(t <= par("usr")[2], t, NA)
    points(x, rep(y, length(t)), pch=c(pch.trans, if(Status | !is.na(aml_date)) pch.end else NA), col=col[c(2,status)])
    par(xpd=xpd)
}
lineStageMF <- function(Date_LF, ERDate, aml_date, Status, y=0, col=1:5, pch.trans=19, pch.end=19, ...){
    xpd <- par("xpd")
    par(xpd=NA)
    t <- as.numeric(Date_LF - ERDate )
    to <- pmin(t, par("usr")[2])
    l <- length(to)
    segments(0, y, to, y, col=col, lend=1, ...)
    status <- 1
    if (is.na(aml_date) & Status==1)
      status <- 2
    else if (!is.na(aml_date))
      status <- 3
    x <- ifelse(t <= par("usr")[2], t, NA)
    points(x, y, pch=c(if(Status | !is.na(aml_date)) pch.end else NA), col=col[status])
    par(xpd=xpd)
}
pastel1 <- brewer.pal(9, "Pastel1")
par(mfrow=c(1,1), mar=c(3,3,1,1), cex=1)
sedimentPlot(-rowMeans(fiveStagePredictedCP[,1:5,], dims=2), y0=1, y1=0,  col=c(pastel1[c(1,2,3,4,5)], "#DDDDDD"))
lines(survfit(os_from_cp ~ 1))
legend('bottomleft', bty='n', col=c("red", "blue", "green", "purple", "orange", "grey"), legend=c("Death in CP", "Death in MF", "Death in AML post CP",
```

```
"Death in AML post MF", "Alive in MF", "Alive in CP"), lty=1)
nStars <- 32; set.seed(42)
os_cp_start <- Surv(cp_data$start.time[wcp], time_death_cp, outcome_from_cp)
#os_cp_start <- Surv(time_death_cp - cp_data$start.time[wcp], outcome_from_cp)

cOs_cp_start <- CoxRFX(cp_data$Z[wcp,], os_cp_start, groups=groups, which.mu = which.mu, nu=1, sigma0=0.01)

s_term <- sample(which(outcome_from_cp==1))
s_cens <- sample(which(outcome_from_cp==0),nStars^2 - length(s_term))
s <- c(s_term, s_cens)
l <- "cOs_cp_start"
t <- os_cp_start
p <- PartialRisk(get(l),  newZ=cp_data$Z[wcp, colnames(get(l)$Z)])
p <- p[,colnames(p)!="Nuisance"]
locations <- hilbertCurve(log2(nStars))+1
mat <- matrix(order(locations[,1], locations[,2]), ncol=nStars)
h <- hclust(dist(t(fiveStagePredictedCP[5*365/25,,s])))#hclust(dist(p[s,]))
o <- h$order # seq_along(h$order)#
##Tiles graph
pdf("tiles.pdf")
layout(mat[nStars:1,])
par(mar=c(0,0,0,0),+.5, bty="n")
for(i in 1:nStars^2){ # Fitted predictions
   sedimentPlot(-fiveStagePredictedCP[seq(1,301,30),1:5,s[o[i]]], x=seq(1,301,30),y0=1, y1=0,  col=c(pastel1[c(1,2,3,4,5)], "#DDDDDD"),
xlab="time",ylab="fraction", xaxt="n", yaxt="n")
   lines(x=seq(1,301,30), y=1-rowSums(fiveStagePredictedCP[seq(1,301,30),1:4,s[o[i]]]), lwd=1)
   i <- s[o[i]]
   lineStageCP(time_mf_cp[i]*365/25, time_death_cp[i]*365/25, cp_data$start.time[wcp][i]*365/25, time_aml_cp[i]*365/25, outcome_from_cp[i],
col=c('#999999','#ddcc00','#990000','#000099','#009900','#990099'), lwd=2, pch.trans=NA, y=0.05)
   #text(x=c(150), y=c(0.2), labels=c(i))
```

```
}
dev.off()
survConcordance(cp_fit$surv ~ predict(cp_fit))
survConcordance(mf_fit$surv ~ predict(mf_fit))
survConcordance(cp_to_mf_fit$surv ~ predict(cp_to_mf_fit))
survConcordance(cp_to_aml_fit$surv ~ predict(cp_to_aml_fit))
survConcordance(mf_to_aml_fit$surv ~ predict(mf_to_aml_fit))
survConcordance(aml_fit$surv ~ predict(aml_fit))
# print(survConcordance(os_cp_start ~ colSums(fiveStagePredictedCP[5*365/25,1:4,])))
# print(survConcordance(os_cp_start ~ colSums(fiveStagePredictedCP[10*365/25,1:4,])))
# print(survConcordance(os_cp_start ~ colSums(fiveStagePredictedCP[15*365/25,1:4,])))
# print(survConcordance(os_cp_start ~ colSums(fiveStagePredictedCP[20*365/25,1:4,])))
# print(survConcordance(os_cp_start ~ colSums(fiveStagePredictedCP[25*365/25,1:4,])))
sink("TGS_output.txt",append=TRUE)
cat("Variance contributions and HRs for each variable")
cbind(colSums(cov(as.matrix(all_data$Z[,names(aml_fit$coeff)])%*%diag(aml_fit$coeff))),exp(aml_fit$coeff))
cbind(colSums(cov(as.matrix(cp_data$Z[,names(cp_fit$coeff)])%*%diag(cp_fit$coeff))),exp(cp_fit$coeff))
cbind(colSums(cov(as.matrix(mf_data$Z[,names(mf_fit$coeff[1:61])])%*%diag(mf_fit$coeff[1:61]))),exp(mf_fit$coeff)[1:61])
cbind(colSums(cov(as.matrix(cp_data$Z[,names(cp_to_mf_fit$coeff)])%*%diag(cp_to_mf_fit$coeff))),exp(cp_to_mf_fit$coeff))
sink()
###Run LOO first, then
TGSZcut_pred<- sapply(lapply(c(1:nrow(TGSZcut)), function(i){
e <- new.env()
if (i %in% (c(1:20)*100))
print(paste("Done",i,"out of",nrow(TGSZcut)))
t <- try(load(paste0("../data/loo/",TGSZcut$UPN[i],".RData"), env=e))
if(class(t)=="try-error") rep(NA, length(times))
else e$fiveStagePredicted[,,1]
}), I, simplify="array")
##Otherwise: TGSZcut_pred<-MultiRFX5(cp_fit, aml_fit, cp_to_mf_fit, mf_fit, aml_fit, TGSZcut, x=365*25)
```

```
newdataplot<-
function(newdata,ET=newdata$ET,PV=newdata$PV,MF=newdata$MF){
pastel2<-c("#BFFFFF","#A3D8D1","#FFBFC9","#FF7387","#D1C299","#A3D8D1","#FF7387")
newdata$MF==MF;newdata$PV==PV;newdata$ET==ET
multistate<-MultiRFX5(cp_fit, aml_fit, cp_to_mf_fit, mf_fit, aml_fit, newdata, x=365*25)
if (MF==0) {
sedimentPlot(-multistate[seq(1,361,30),1:5,1], x=seq(1,361,30),y0=1, y1=0,  col=c(pastel2[c(1,2,3,4,5)], "#D6E3DE"), xlab="Time from diagnosis
(years)",ylab="Proportion of patients", xaxt="n",yaxt="n")
   lines(x=seq(1,361,30), y=1-rowSums(multistate[seq(1,361,30),1:4,1]), lwd=1)
segments(x0=0,y0=0.5,x1=newdata$Death/25,y1=0.5)
points(x=newdata$Death/25,y=0.5,cex=newdata$DeathC,pch=19);points(x=newdata$Death/25,y=0.5,cex=newdata$DeathC*0.85,pch=19,col="#BFFFFF")
if(newdata$DeathC==1&newdata$MFTC==1&newdata$AMLTC==0){points(x=newdata$Death/25,y=0.5,cex=0.9,pch=19,col="#A3D8D1")}
if(newdata$DeathC==1&newdata$MFTC==1&newdata$AMLTC==1){points(x=newdata$Death/25,y=0.5,cex=0.9,pch=19,col="#FF7387")}
if(newdata$DeathC==1&newdata$MFTC==0&newdata$AMLTC==1){points(x=newdata$Death/25,y=0.5,cex=0.9,pch=19,col="#FFBFC9")}
points(x=newdata$MFT/25,y=0.5,cex=newdata$MFTC*0.5,pch=18)
} else {
sedimentPlot(-multistate[seq(1,361,30),6:7,1], x=seq(1,361,30),y0=1, y1=0,  col=c(pastel2[c(6,7)], "#D1C299"), xlab="Time from diagnosis
(years)",ylab="Proportion of patients", xaxt="n",yaxt="n")
   lines(x=seq(1,361,30), y=1-rowSums(multistate[seq(1,361,30),6:7,1]), lwd=1)
segments(x0=0,y0=0.5,x1=newdata$Death/25,y1=0.5);points(x=newdata$Death/25,y=0.5,cex=newdata$DeathC,pch=19);points(x=newdata$Death/25,y=0.5
,cex=newdata$AMLTC*0.85,pch=19,col="#FF7387")
}
}
newdataplotLOO<-
function(newdata,multistate,i){
pastel2<-c("#BFFFFF","#A3D8D1","#FFBFC9","#FF7387","#D1C299","#A3D8D1","#FF7387")
newdata<-newdata[i,];MF<-newdata$MF;PV<-newdata$PV;ET<-newdata$ET
if (MF==0) {
sedimentPlot(-multistate[seq(1,361,30),1:5,i], x=seq(1,361,30),y0=1, y1=0,  col=c(pastel2[c(1,2,3,4,5)], "#D6E3DE"), xlab="Time from diagnosis
(years)",ylab="Proportion of patients", xaxt="n",yaxt="n")
```

```
   lines(x=seq(1,361,30), y=1-rowSums(multistate[seq(1,361,30),1:4,i]), lwd=1)
segments(x0=0,y0=0.5,x1=newdata$Death/25,y1=0.5)
points(x=newdata$Death/25,y=0.5,cex=newdata$DeathC*2,pch=19);points(x=newdata$Death/25,y=0.5,cex=newdata$DeathC*1.8,pch=19,col="#BFFFFF")
if(newdata$DeathC==1&newdata$MFTC==1&newdata$AMLTC==0){points(x=newdata$Death/25,y=0.5,cex=1.8,pch=19,col="#A3D8D1")}
if(newdata$DeathC==1&newdata$MFTC==1&newdata$AMLTC==1){points(x=newdata$Death/25,y=0.5,cex=1.8,pch=19,col="#FF7387")}
if(newdata$DeathC==1&newdata$MFTC==0&newdata$AMLTC==1){points(x=newdata$Death/25,y=0.5,cex=1.8,pch=19,col="#FFBFC9")}
points(x=newdata$MFT/25,y=0.5,cex=newdata$MFTC,pch=4)
} else {
sedimentPlot(-multistate[seq(1,361,30),6:7,i], x=seq(1,361,30),y0=1, y1=0,  col=c(pastel2[c(6,7)], "#D1C299"), xlab="Time from diagnosis
(years)",ylab="Proportion of patients", xaxt="n",yaxt="n")
   lines(x=seq(1,361,30), y=1-rowSums(multistate[seq(1,361,30),6:7,i]), lwd=1)
segments(x0=0,y0=0.5,x1=newdata$Death/25,y1=0.5);points(x=newdata$Death/25,y=0.5,cex=newdata$DeathC*2,pch=19);points(x=newdata$Death/25,y=
0.5,cex=newdata$DeathC*1.8,pch=19,col="#A3D8D1");points(x=newdata$Death/25,y=0.5,cex=newdata$AMLTC*1.8,pch=19,col="#FF7387")
}
}
ape2<-function (x, time,censor, timepoint, censored = "conditional")
{
   surv<-Surv(time,censor)
   status <- survStatus(surv, timepoint, censored = censored)
   p<-length(which(status==1))/length(censor);unc<-p*(1-p)
   err <- c(unc=unc, brier = mean((x-status)^2, na.rm = TRUE),brier0=mean((rep(0,length(censor))-status)^2,na.rm=TRUE),
brier1=mean((rep(1,length(censor))-status)^2,na.rm=TRUE), brierp=mean((rep(p,length(censor))-status)^2,na.rm=TRUE), abs = mean(abs(x-status),
      na.rm = TRUE))
   return(err)
}
sink("TGS_output.txt",append=TRUE)
cat("Concordances for full multistate model")
for(i in c(5,10,15,20)){
cat("OS concordance for CP, using",i, "yr predictions: Concordance:",
survConcordance(Surv(TGSZcut$Death,TGSZcut$DeathC)[which(TGSZcut$MF==0)]~colSums(TGSZcut_pred[i*365/25,1:4,which(TGSZcut$MF==0)]))$
```

```
concordance, " stderr: ",
survConcordance(Surv(TGSZcut$Death,TGSZcut$DeathC)[which(TGSZcut$MF==0)]~colSums(TGSZcut_pred[i*365/25,1:4,which(TGSZcut$MF==0)]))$
std.err,"\n")
}
for(i in c(5,10,15,20)){
cat("MF transformation concordance for CP, using",i, "yr predictions: Concordance:",
survConcordance(Surv(TGSZcut$MFT,TGSZcut$MFTC)[which(TGSZcut$MF==0)]~colSums(TGSZcut_pred[i*365/25,c(2,4,5),which(TGSZcut$MF==0)]
))$concordance, " stderr: ",
survConcordance(Surv(TGSZcut$MFT,TGSZcut$MFTC)[which(TGSZcut$MF==0)]~colSums(TGSZcut_pred[i*365/25,c(2,4,5),which(TGSZcut$MF==0)]
))$std.err,"\n")
}
for(i in c(5,10,15,20)){
cat("AML transformation concordance for CP, using",i, "yr predictions: Concordance:",
survConcordance(Surv(TGSZcut$AMLT,TGSZcut$AMLTC)[which(TGSZcut$MF==0)]~colSums(TGSZcut_pred[i*365/25,c(3,4),which(TGSZcut$MF==0
)]))$concordance, " stderr: ",
survConcordance(Surv(TGSZcut$AMLT,TGSZcut$AMLTC)[which(TGSZcut$MF==0)]~colSums(TGSZcut_pred[i*365/25,c(3,4),which(TGSZcut$MF==0
)]))$std.err,"\n")
}
for(i in c(5,10,15,20)){
cat("EFS concordance for CP, using",i, "yr predictions: Concordance:",
survConcordance(Surv(TGSZcut$EFS,TGSZcut$EFSC)[which(TGSZcut$MF==0)]~colSums(TGSZcut_pred[i*365/25,1:5,which(TGSZcut$MF==0)]))$co
ncordance, " stderr: ",
survConcordance(Surv(TGSZcut$EFS,TGSZcut$EFSC)[which(TGSZcut$MF==0)]~colSums(TGSZcut_pred[i*365/25,1:5,which(TGSZcut$MF==0)]))$std.
err,"\n")
}

for(i in c(5,10,15,20)){
cat("OS concordance for ET, using",i, "yr predictions: Concordance:",
survConcordance(Surv(TGSZcut$Death,TGSZcut$DeathC)[which(TGSZcut$ET==1)]~colSums(TGSZcut_pred[i*365/25,1:4,which(TGSZcut$ET==1)]))$c
oncordance, " stderr: ",
```

```
survConcordance(Surv(TGSZcut$Death,TGSZcut$DeathC)[which(TGSZcut$ET==1)]~colSums(TGSZcut_pred[i*365/25,1:4,which(TGSZcut$ET==1)]))$s
td.err,"\n")
}
for(i in c(5,10,15,20)){
cat("MF transformation concordance for ET, using",i, "yr predictions: Concordance:",
survConcordance(Surv(TGSZcut$MFT,TGSZcut$MFTC)[which(TGSZcut$ET==1)]~colSums(TGSZcut_pred[i*365/25,c(2,4,5),which(TGSZcut$ET==1)])
)$concordance, " stderr: ",
survConcordance(Surv(TGSZcut$MFT,TGSZcut$MFTC)[which(TGSZcut$ET==1)]~colSums(TGSZcut_pred[i*365/25,c(2,4,5),which(TGSZcut$ET==1)])
)$std.err,"\n")
}
for(i in c(5,10,15,20)){
cat("AML transformation concordance for ET, using",i, "yr predictions: Concordance:",
survConcordance(Surv(TGSZcut$AMLT,TGSZcut$AMLTC)[which(TGSZcut$ET==1)]~colSums(TGSZcut_pred[i*365/25,c(3,4),which(TGSZcut$ET==1)
]))$concordance, " stderr: ",
survConcordance(Surv(TGSZcut$AMLT,TGSZcut$AMLTC)[which(TGSZcut$ET==1)]~colSums(TGSZcut_pred[i*365/25,c(3,4),which(TGSZcut$ET==1)
]))$std.err,"\n")
}
for(i in c(5,10,15,20)){
cat("EFS concordance for ET, using",i, "yr predictions: Concordance:",
survConcordance(Surv(TGSZcut$EFS,TGSZcut$EFSC)[which(TGSZcut$ET==1)]~colSums(TGSZcut_pred[i*365/25,1:5,which(TGSZcut$ET==1)]))$con
cordance, " stderr: ",
survConcordance(Surv(TGSZcut$EFS,TGSZcut$EFSC)[which(TGSZcut$ET==1)]~colSums(TGSZcut_pred[i*365/25,1:5,which(TGSZcut$ET==1)]))$std.e
rr,"\n")
}

for(i in c(5,10,15,20)){
cat("EFS concordance for MF, using",i, "yr predictions: Concordance:",
survConcordance(Surv(TGSZcut$EFS,TGSZcut$EFSC)[which(TGSZcut$MF==1)]~colSums(TGSZcut_pred[i*365/25,6:7,which(TGSZcut$MF==1)]))$co
ncordance, " stderr: ",
survConcordance(Surv(TGSZcut$EFS,TGSZcut$EFSC)[which(TGSZcut$MF==1)]~colSums(TGSZcut_pred[i*365/25,6:7,which(TGSZcut$MF==1)]))$std.
```

```
err,"\n")
}
for(i in c(5,10,15,20)){
cat("AML concordance for MF, using",i, "yr predictions: Concordance:",
survConcordance(Surv(TGSZcut$AMLT,TGSZcut$AMLTC)[which(TGSZcut$MF==1)]~TGSZcut_pred[i*365/25,7,which(TGSZcut$MF==1)])$concordan
ce, " stderr: ",
survConcordance(Surv(TGSZcut$AMLT,TGSZcut$AMLTC)[which(TGSZcut$MF==1)]~TGSZcut_pred[i*365/25,7,which(TGSZcut$MF==1)])$std.err,"\n"
)
}
cat("Prediction uncertainty, Brier score, APE for full multistate model")
for(i in c(5,10,15,20)){
cat("OS in CP at",i,"yrs",
ape2(1-
colSums(TGSZcut_pred[i*365/25,1:4,which(TGSZcut$MF==0)]),TGSZcut$Death[which(TGSZcut$MF==0)],TGSZcut$DeathC[which(TGSZcut$MF==0)]
,i*365.25),"\n")
}
for(i in c(5,10,15,20)){
cat("MF transformation in CP at",i,"yrs",
ape2(1-
colSums(TGSZcut_pred[i*365/25,c(2,4,5),which(TGSZcut$MF==0)]),TGSZcut$MFT[which(TGSZcut$MF==0)],TGSZcut$MFTC[which(TGSZcut$MF==
0)],i*365.25),"\n")
}
for(i in c(5,10,15,20)){
cat("AML transformation in CP at",i,"yrs",
ape2(1-
colSums(TGSZcut_pred[i*365/25,c(3,4),which(TGSZcut$MF==0)]),TGSZcut$AMLT[which(TGSZcut$MF==0)],TGSZcut$AMLTC[which(TGSZcut$MF
==0)],i*365.25),"\n")
}
for(i in c(5,10,15,20)){
cat("EFS in CP at",i,"yrs",
```

```
ape2(1-
colSums(TGSZcut_pred[i*365/25,1:5,which(TGSZcut$MF==0)]),TGSZcut$EFS[which(TGSZcut$MF==0)],TGSZcut$EFSC[which(TGSZcut$MF==0)],i*
365.25),"\n")
}

cat("Prediction uncertainty, Brier score, APE for full multistate model")
for(i in c(5,10,15,20)){
cat("OS in ET at",i,"yrs",
ape2(1-
colSums(TGSZcut_pred[i*365/25,1:4,which(TGSZcut$ET==1)]),TGSZcut$Death[which(TGSZcut$ET==1)],TGSZcut$DeathC[which(TGSZcut$ET==1)],i
*365.25),"\n")
}
for(i in c(5,10,15,20)){
cat("MF transformation in ET at",i,"yrs",
ape2(1-
colSums(TGSZcut_pred[i*365/25,c(2,4,5),which(TGSZcut$ET==1)]),TGSZcut$MFT[which(TGSZcut$ET==1)],TGSZcut$MFTC[which(TGSZcut$ET==1
)],i*365.25),"\n")
}
for(i in c(5,10,15,20)){
cat("AML transformation in ET at",i,"yrs",
ape2(1-
colSums(TGSZcut_pred[i*365/25,c(3,4),which(TGSZcut$ET==1)]),TGSZcut$AMLT[which(TGSZcut$ET==1)],TGSZcut$AMLTC[which(TGSZcut$ET==
1)],i*365.25),"\n")
}
for(i in c(5,10,15,20)){
cat("EFS in ET at",i,"yrs",
ape2(1-
colSums(TGSZcut_pred[i*365/25,1:5,which(TGSZcut$ET==1)]),TGSZcut$EFS[which(TGSZcut$ET==1)],TGSZcut$EFSC[which(TGSZcut$ET==1)],i*36
5.25),"\n")
}
```

```
for(i in c(5,10,15,20)){
cat("EFS in MF at",i,"yrs",
ape2(1-
colSums(TGSZcut_pred[i*365/25,6:7,which(TGSZcut$MF==1)]),TGSZcut$EFS[which(TGSZcut$MF==1)],TGSZcut$EFSC[which(TGSZcut$MF==1)],i*
365.25),"\n")
}
for(i in c(5,10,15,20)){
cat("AML transformation in MF at",i,"yrs",
ape2(1-
TGSZcut_pred[i*365/25,7,which(TGSZcut$MF==1)],TGSZcut$AMLT[which(TGSZcut$MF==1)],TGSZcut$AMLTC[which(TGSZcut$MF==1)],i*365.25)
,"\n")
}

cat("Other models:")
cat("OS concordance for CP, using HMR: Concordance:",
survConcordance(Surv(TGSZcut$Death,TGSZcut$DeathC)[which(TGSZcut$MF==0)]~TGSZcut$HMR[which(TGSZcut$MF==0)])$concordance, " stderr:
", survConcordance(Surv(TGSZcut$Death,TGSZcut$DeathC)[which(TGSZcut$MF==0)]~TGSZcut$HMR[which(TGSZcut$MF==0)])$std.err,"\n")
cat("MF transformation concordance for HMR: Concordance:",
survConcordance(Surv(TGSZcut$MFT,TGSZcut$MFTC)[which(TGSZcut$MF==0)]~TGSZcut$HMR[which(TGSZcut$MF==0)])$concordance, " stderr: ",
survConcordance(Surv(TGSZcut$MFT,TGSZcut$MFTC)[which(TGSZcut$MF==0)]~TGSZcut$HMR[which(TGSZcut$MF==0)])$std.err,"\n")
cat("AML transformation concordance for HMR: Concordance:",
survConcordance(Surv(TGSZcut$AMLT,TGSZcut$AMLTC)[which(TGSZcut$MF==0)]~TGSZcut$HMR[which(TGSZcut$MF==0)])$concordance, "
stderr: ", survConcordance(Surv(TGSZcut$AMLT,TGSZcut$AMLTC)[which(TGSZcut$MF==0)]~TGSZcut$HMR[which(TGSZcut$MF==0)])$std.err,"\n")
cat("EFS concordance for CP, using HMR. Concordance:",
survConcordance(Surv(TGSZcut$EFS,TGSZcut$EFSC)[which(TGSZcut$MF==0)]~TGSZcut$HMR[which(TGSZcut$MF==0)])$concordance, " stderr: ",
survConcordance(Surv(TGSZcut$EFS,TGSZcut$EFSC)[which(TGSZcut$MF==0)]~TGSZcut$HMR[which(TGSZcut$MF==0)])$std.err,"\n")

cat("OS concordance for CP, using Age/thrombosis: Concordance:",
survConcordance(Surv(TGSZcut$Death,TGSZcut$DeathC)[which(TGSZcut$MF==0)]~TGSZcut$PT60[which(TGSZcut$MF==0)])$concordance, " stderr:
```

```
", survConcordance(Surv(TGSZcut$Death,TGSZcut$DeathC)[which(TGSZcut$MF==0)]~TGSZcut$PT60[which(TGSZcut$MF==0)])$std.err,"\n")
cat("MF transformation concordance for Age/Thrombosis: Concordance:",
survConcordance(Surv(TGSZcut$MFT,TGSZcut$MFTC)[which(TGSZcut$MF==0)]~TGSZcut$PT60[which(TGSZcut$MF==0)])$concordance, " stderr: ",
survConcordance(Surv(TGSZcut$MFT,TGSZcut$MFTC)[which(TGSZcut$MF==0)]~TGSZcut$PT60[which(TGSZcut$MF==0)])$std.err,"\n")
cat("AML transformation concordance for Age/thrombosis: Concordance:",
survConcordance(Surv(TGSZcut$AMLT,TGSZcut$AMLTC)[which(TGSZcut$MF==0)]~TGSZcut$PT60[which(TGSZcut$MF==0)])$concordance, "
stderr: ", survConcordance(Surv(TGSZcut$AMLT,TGSZcut$AMLTC)[which(TGSZcut$MF==0)]~TGSZcut$PT60[which(TGSZcut$MF==0)])$std.err,"\n")
cat("EFS concordance for CP, using Age/Thrombosis: Concordance:",
survConcordance(Surv(TGSZcut$EFS,TGSZcut$EFSC)[which(TGSZcut$MF==0)]~TGSZcut$PT60[which(TGSZcut$MF==0)])$concordance, " stderr: ",
survConcordance(Surv(TGSZcut$EFS,TGSZcut$EFSC)[which(TGSZcut$MF==0)]~TGSZcut$PT60[which(TGSZcut$MF==0)])$std.err,"\n")

cat("OS concordance for CP, using IPSET: Concordance:",
survConcordance(Surv(TGSZcut$Death,TGSZcut$DeathC)[which(TGSZcut$ET==1)]~TGSZcut$IPSET[which(TGSZcut$ET==1)])$concordance, " stderr:
", survConcordance(Surv(TGSZcut$Death,TGSZcut$DeathC)[which(TGSZcut$ET==1)]~TGSZcut$IPSET[which(TGSZcut$ET==1)])$std.err,"\n")
cat("MF transformation concordance for IPSET: Concordance:",
survConcordance(Surv(TGSZcut$MFT,TGSZcut$MFTC)[which(TGSZcut$ET==1)]~TGSZcut$IPSET[which(TGSZcut$ET==1)])$concordance, " stderr: ",
survConcordance(Surv(TGSZcut$MFT,TGSZcut$MFTC)[which(TGSZcut$ET==1)]~TGSZcut$IPSET[which(TGSZcut$ET==1)])$std.err,"\n")
cat("AML transformation concordance for IPSET: Concordance:",
survConcordance(Surv(TGSZcut$AMLT,TGSZcut$AMLTC)[which(TGSZcut$ET==1)]~TGSZcut$IPSET[which(TGSZcut$ET==1)])$concordance, "
stderr: ", survConcordance(Surv(TGSZcut$AMLT,TGSZcut$AMLTC)[which(TGSZcut$ET==1)]~TGSZcut$IPSET[which(TGSZcut$ET==1)])$std.err,"\n")
cat("EFS concordance for CP, using IPSET. Concordance:",
survConcordance(Surv(TGSZcut$EFS,TGSZcut$EFSC)[which(TGSZcut$ET==1)]~TGSZcut$IPSET[which(TGSZcut$ET==1)])$concordance, " stderr: ",
survConcordance(Surv(TGSZcut$EFS,TGSZcut$EFSC)[which(TGSZcut$ET==1)]~TGSZcut$IPSET[which(TGSZcut$ET==1)])$std.err,"\n")

cat("OS concordance for CP, using CALRASXL1: Concordance:",
survConcordance(Surv(TGSZcut$Death,TGSZcut$DeathC)[which(TGSZcut$MF==0)]~TGSZcut$CALRASXL1[which(TGSZcut$MF==0)])$concordance,
" stderr: ",
survConcordance(Surv(TGSZcut$Death,TGSZcut$DeathC)[which(TGSZcut$MF==0)]~TGSZcut$CALRASXL1[which(TGSZcut$MF==0)])$std.err,"\n")
cat("MF transformation concordance in CP for CALRASXL1: Concordance:",
survConcordance(Surv(TGSZcut$MFT,TGSZcut$MFTC)[which(TGSZcut$MF==0)]~TGSZcut$CALRASXL1[which(TGSZcut$MF==0)])$concordance, "
```

```
stderr: ",
survConcordance(Surv(TGSZcut$MFT,TGSZcut$MFTC)[which(TGSZcut$MF==0)]~TGSZcut$CALRASXL1[which(TGSZcut$MF==0)])$std.err,"\n")
cat("AML transformation concordance in CP for CALRASXL1: Concordance:",
survConcordance(Surv(TGSZcut$AMLT,TGSZcut$AMLTC)[which(TGSZcut$MF==0)]~TGSZcut$CALRASXL1[which(TGSZcut$MF==0)])$concordanc
e, " stderr: ",
survConcordance(Surv(TGSZcut$AMLT,TGSZcut$AMLTC)[which(TGSZcut$MF==0)]~TGSZcut$CALRASXL1[which(TGSZcut$MF==0)])$std.err,"\n")
cat("EFS concordance for CP, using CALRASXL1. Concordance:",
survConcordance(Surv(TGSZcut$EFS,TGSZcut$EFSC)[which(TGSZcut$MF==0)]~TGSZcut$CALRASXL1[which(TGSZcut$MF==0)])$concordance, "
stderr: ",
survConcordance(Surv(TGSZcut$EFS,TGSZcut$EFSC)[which(TGSZcut$MF==0)]~TGSZcut$CALRASXL1[which(TGSZcut$MF==0)])$std.err,"\n")

cat("EFS concordance for HMR in MF: ",
survConcordance(Surv(TGSZcut$Death,TGSZcut$DeathC)[which(TGSZcut$MF==1)]~TGSZcut$HMR[which(TGSZcut$MF==1)])$concordance, " stderr:
", survConcordance(Surv(TGSZcut$Death,TGSZcut$DeathC)[which(TGSZcut$MF==1)]~TGSZcut$HMR[which(TGSZcut$MF==1)])$std.err,"\n")
cat("AML transformation concordance for HMR in MF: ",
survConcordance(Surv(TGSZcut$AMLT,TGSZcut$AMLTC)[which(TGSZcut$MF==1)]~TGSZcut$HMR[which(TGSZcut$MF==1)])$concordance, "
stderr: ", survConcordance(Surv(TGSZcut$AMLT,TGSZcut$AMLTC)[which(TGSZcut$MF==1)]~TGSZcut$HMR[which(TGSZcut$MF==1)])$std.err,"\n")

cat("EFS concordance for PT60 in MF: ",
survConcordance(Surv(TGSZcut$Death,TGSZcut$DeathC)[which(TGSZcut$MF==1)]~TGSZcut$PT60[which(TGSZcut$MF==1)])$concordance, " stderr:
", survConcordance(Surv(TGSZcut$Death,TGSZcut$DeathC)[which(TGSZcut$MF==1)]~TGSZcut$PT60[which(TGSZcut$MF==1)])$std.err,"\n")
cat("AML transformation concordance for PT60 in MF: ",
survConcordance(Surv(TGSZcut$AMLT,TGSZcut$AMLTC)[which(TGSZcut$MF==1)]~TGSZcut$PT60[which(TGSZcut$MF==1)])$concordance, "
stderr: ", survConcordance(Surv(TGSZcut$AMLT,TGSZcut$AMLTC)[which(TGSZcut$MF==1)]~TGSZcut$PT60[which(TGSZcut$MF==1)])$std.err,"\n")

cat("EFS concordance for CALRASXL1 in MF: ",
survConcordance(Surv(TGSZcut$Death,TGSZcut$DeathC)[which(TGSZcut$MF==1)]~TGSZcut$CALRASXL1[which(TGSZcut$MF==1)])$concordance,
" stderr: ",
survConcordance(Surv(TGSZcut$Death,TGSZcut$DeathC)[which(TGSZcut$MF==1)]~TGSZcut$CALRASXL1[which(TGSZcut$MF==1)])$std.err,"\n")
cat("AML transformation concordance for CALRASXL1 in MF: ",
```

```
survConcordance(Surv(TGSZcut$AMLT,TGSZcut$AMLTC)[which(TGSZcut$MF==1)]~TGSZcut$CALRASXL1[which(TGSZcut$MF==1)])$concordanc
e, " stderr: ",
survConcordance(Surv(TGSZcut$AMLT,TGSZcut$AMLTC)[which(TGSZcut$MF==1)]~TGSZcut$CALRASXL1[which(TGSZcut$MF==1)])$std.err,"\n")

for(i in c(5,10,15,20)){
print(cbind(aggregate(1-colSums(TGSZcut_pred[i*365/25,1:5,which(TGSZcut$MF==0)]),by=list(quantileCut(1-
colSums(TGSZcut_pred[i*365/25,1:5,which(TGSZcut$MF==0)]),20)),FUN=median),summary(survfit(Surv(EFS/365.25,EFSC)[which(TGSZcut$MF==0)]
~quantileCut(1-
colSums(TGSZcut_pred[i*365/25,1:5,which(TGSZcut$MF==0)]),20),data=TGSZcut),times=c(i),extend=TRUE)$surv,summary(survfit(Surv(EFS/365.25,E
FSC)[which(TGSZcut$MF==0)]~quantileCut(1-
colSums(TGSZcut_pred[i*365/25,1:5,which(TGSZcut$MF==0)]),20),data=TGSZcut),times=c(i),extend=TRUE)$std.err))
}

for(i in c(5,10,15,20)){
print(cbind(aggregate(1-colSums(TGSZcut_pred[i*365/25,6:7,which(TGSZcut$MF==1)]),by=list(quantileCut(1-
colSums(TGSZcut_pred[i*365/25,6:7,which(TGSZcut$MF==1)]),15)),FUN=median),summary(survfit(Surv(Death/365.25,DeathC)[which(TGSZcut$MF==
1)]~quantileCut(1-
colSums(TGSZcut_pred[i*365/25,6:7,which(TGSZcut$MF==1)]),15),data=TGSZcut),times=c(i),extend=TRUE)$surv,summary(survfit(Surv(Death/365.25,
DeathC)[which(TGSZcut$MF==1)]~quantileCut(1-
colSums(TGSZcut_pred[i*365/25,6:7,which(TGSZcut$MF==1)]),15),data=TGSZcut),times=c(i),extend=TRUE)$std.err))
}



sink()

## External validation cohort
florMF<-read.table("florMF2.csv",sep=",",header=TRUE)
florMF$DeathC[which(florMF$AMLTC==1)]<-1
florMF_pred<-MultiRFX5(cp_fit, aml_fit, cp_to_mf_fit, mf_fit, aml_fit, florMF, x=365*25)
```

```
sink("TGS_output.txt",append=TRUE)
cat("Analysis of external MF cohort")
for(i in c(5,10,15,20)){
cat("EFS concordance for MF, using",i, "yr predictions: Concordance:",
survConcordance(Surv(florMF$Death,florMF$DeathC)[which(florMF$MF==1)]~colSums(florMF_pred[i*365/25,6:7,which(florMF$MF==1)]))$concordan
ce, " stderr: ",
survConcordance(Surv(florMF$Death,florMF$DeathC)[which(florMF$MF==1)]~colSums(florMF_pred[i*365/25,6:7,which(florMF$MF==1)]))$std.err,"\n
")
}
for(i in c(5,10,15,20)){
cat("AML concordance for MF, using",i, "yr predictions: Concordance:",
survConcordance(Surv(florMF$AMLT,florMF$AMLTC)[which(florMF$MF==1)]~florMF_pred[i*365/25,7,which(florMF$MF==1)])$concordance, "
stderr: ", survConcordance(Surv(florMF$AMLT,florMF$AMLTC)[which(florMF$MF==1)]~florMF_pred[i*365/25,7,which(florMF$MF==1)])$std.err,"\n")
}
for(i in c(5,10,15,20)){
cat("EFS in MF at",i,"yrs",
ape2(1-
colSums(florMF_pred[i*365/25,6:7,which(florMF$MF==1)]),florMF$Death[which(florMF$MF==1)],florMF$DeathC[which(florMF$MF==1)],i*365.25),"
\n")
}
for(i in c(5,10,15,20)){
cat("AML transformation in MF at",i,"yrs",
ape2(1-
florMF_pred[i*365/25,7,which(florMF$MF==1)],florMF$AMLT[which(florMF$MF==1)],florMF$AMLTC[which(florMF$MF==1)],i*365.25),"\n")
}

cat("EFS concordance for MF using DIPSS: Concordance:",
survConcordance(Surv(florMF$Death,florMF$DeathC)[which(florMF$MF==1)]~florMF$DIPSS)$concordance, " stderr: ",
survConcordance(Surv(florMF$Death,florMF$DeathC)[which(florMF$MF==1)]~florMF$DIPSS)$std.err,"\n")
```

```
cat("AML concordance for MF using DIPSS: Concordance:",
survConcordance(Surv(florMF$AMLT,florMF$AMLTC)[which(florMF$MF==1)]~florMF$DIPSS)$concordance, " stderr: ",
survConcordance(Surv(florMF$AMLT,florMF$AMLTC)[which(florMF$MF==1)]~florMF$DIPSS)$std.err,"\n")

cat("EFS concordance for MF using IPSS: Concordance:",
survConcordance(Surv(florMF$Death,florMF$DeathC)[which(florMF$MF==1)]~florMF$IPSS)$concordance, " stderr: ",
survConcordance(Surv(florMF$Death,florMF$DeathC)[which(florMF$MF==1)]~florMF$IPSS)$std.err,"\n")

cat("AML concordance for MF using IPSS: Concordance:",
survConcordance(Surv(florMF$AMLT,florMF$AMLTC)[which(florMF$MF==1)]~florMF$IPSS)$concordance, " stderr: ",
survConcordance(Surv(florMF$AMLT,florMF$AMLTC)[which(florMF$MF==1)]~florMF$IPSS)$std.err,"\n")

cat("EFS concordance for MF using HMR: Concordance:",
survConcordance(Surv(florMF$Death,florMF$DeathC)[which(florMF$MF==1)]~florMF$HMR)$concordance, " stderr: ",
survConcordance(Surv(florMF$Death,florMF$DeathC)[which(florMF$MF==1)]~florMF$HMR)$std.err,"\n")

cat("AML concordance for MF using HMR: Concordance:",
survConcordance(Surv(florMF$AMLT,florMF$AMLTC)[which(florMF$MF==1)]~florMF$HMR)$concordance, " stderr: ",
survConcordance(Surv(florMF$AMLT,florMF$AMLTC)[which(florMF$MF==1)]~florMF$HMR)$std.err,"\n")

for(i in c(5,10,15,20)){
print(cbind(aggregate(1-colSums(florMF_pred[i*365/25,6:7,which(florMF$MF==1)]),by=list(quantileCut(1-
colSums(florMF_pred[i*365/25,6:7,which(florMF$MF==1)]),10)),FUN=median),summary(survfit(Surv(Death/365.25,DeathC)[which(florMF$MF==1)]~q
uantileCut(1-
colSums(florMF_pred[i*365/25,6:7,which(florMF$MF==1)]),10),data=florMF),times=c(i),extend=TRUE)$surv,summary(survfit(Surv(Death/365.25,Death
C)[which(florMF$MF==1)]~quantileCut(1-
colSums(florMF_pred[i*365/25,6:7,which(florMF$MF==1)]),10),data=florMF),times=c(i),extend=TRUE)$std.err))
}
```

```
cat("\n External CP cohort")
florCP<-read.table("florCP_backup.csv",sep=",",header=TRUE)
florCP<-florCP[which(florCP$OLD==0),]
florCP$DeathC[which(florCP$AMLTC==1)]<-1
florCP_pred<-MultiRFX5(cp_fit, aml_fit, cp_to_mf_fit, mf_fit, aml_fit, florCP, x=365*25)


for(i in c(5,10,15,20)){
cat("OS concordance for CP, using",i, "yr predictions: Concordance:",
survConcordance(Surv(florCP$Death,florCP$DeathC)~colSums(florCP_pred[i*365/25,1:4,]))$concordance, " stderr: ",
survConcordance(Surv(florCP$Death,florCP$DeathC)~colSums(florCP_pred[i*365/25,1:4,]))$std.err,"\n")
}
for(i in c(5,10,15,20)){
cat("AML concordance for CP, using",i, "yr predictions: Concordance:",
survConcordance(Surv(florCP$AMLT,florCP$AMLTC)~colSums(florCP_pred[i*365/25,3:4,]))$concordance, " stderr: ",
survConcordance(Surv(florCP$AMLT,florCP$AMLTC)~colSums(florCP_pred[i*365/25,3:4,]))$std.err,"\n")
}

for(i in c(5,10,15,20)){
cat("MFT concordance for CP, using",i, "yr predictions: Concordance:",
survConcordance(Surv(florCP$MFT,florCP$MFTC)~colSums(florCP_pred[i*365/25,c(2,4,5),]))$concordance, " stderr: ",
survConcordance(Surv(florCP$MFT,florCP$MFTC)~colSums(florCP_pred[i*365/25,c(2,4,5),]))$std.err,"\n")
}
for(i in c(5,10,15,20)){
cat("EFS concordance for CP, using",i, "yr predictions: Concordance:",
survConcordance(Surv(florCP$EFS,florCP$EFSC)~colSums(florCP_pred[i*365/25,1:5,]))$concordance, " stderr: ",
survConcordance(Surv(florCP$EFS,florCP$EFSC)~colSums(florCP_pred[i*365/25,1:5,]))$std.err,"\n")
}
```

```
for(i in c(5,10,15,20)){
print(cbind(aggregate(1-colSums(florCP_pred[i*365/25,1:5,]),by=list(quantileCut(1-
colSums(florCP_pred[i*365/25,1:5,]),10)),FUN=median),summary(survfit(Surv(EFS/365.25,EFSC)~quantileCut(1-
colSums(florCP_pred[i*365/25,1:5,]),10),data=florCP),times=c(i),extend=TRUE)$surv,summary(survfit(Surv(EFS/365.25,EFSC)~quantileCut(1-
colSums(florCP_pred[i*365/25,1:5,]),10),data=florCP),times=c(i),extend=TRUE)$std.err))
}

for(i in c(5,10,15,20)){
cat("Unc/Briar/Ape. OS in CP at",i,"yrs",
ape2(1-colSums(florCP_pred[i*365/25,1:4,]),florCP$Death,florCP$DeathC,i*365.25),"\n")
}
for(i in c(5,10,15,20)){
cat("Unc/Briar/Ape. AMLT in CP at",i,"yrs",
ape2(1-colSums(florCP_pred[i*365/25,3:4,]),florCP$AMLT,florCP$AMLTC,i*365.25),"\n")
}
for(i in c(5,10,15,20)){
cat("Unc/Briar/Ape. MFT in CP at",i,"yrs",
ape2(1-colSums(florCP_pred[i*365/25,c(2,4,5),]),florCP$MFT,florCP$MFTC,i*365.25),"\n")
}
for(i in c(5,10,15,20)){
cat("Unc/Briar/Ape. EFS in CP at",i,"yrs",
ape2(1-colSums(florCP_pred[i*365/25,1:5,]),florCP$EFS,florCP$EFSC,i*365.25),"\n")
}



##AUC.uno

for(i in c(5,10,15,20)){
cat("Uno. OS in CP at",i,"yrs",
```

```
AUC.uno(Surv(TGSZcut$Death,TGSZcut$DeathC)[which(TGSZcut$MF==0&TGSZcut$DeathC>=0)],Surv(TGSZcut$Death,TGSZcut$DeathC)[which(TG
SZcut$MF==0&TGSZcut$DeathC>=0)],colSums(TGSZcut_pred[i*365/25,1:4,which(TGSZcut$MF==0&TGSZcut$DeathC>=0)]),seq(100,365*25,50))$iau
c, "\n")
}

for(i in c(5,10,15,20)){
cat("Uno. MFT in CP at",i,"yrs",
AUC.uno(Surv(TGSZcut$MFT,TGSZcut$MFTC)[which(TGSZcut$MF==0&TGSZcut$MFTC>=0)],Surv(TGSZcut$MFT,TGSZcut$MFTC)[which(TGSZc
ut$MF==0&TGSZcut$MFTC>=0)],colSums(TGSZcut_pred[i*365/25,c(2,4,5),which(TGSZcut$MF==0&TGSZcut$MFTC>=0)]),seq(100,365*25,50))$iauc
, "\n")
}

for(i in c(5,10,15,20)){
cat("Uno. AML in CP at",i,"yrs",
AUC.uno(Surv(TGSZcut$AMLT,TGSZcut$AMLTC)[which(TGSZcut$MF==0&TGSZcut$AMLTC>=0)],Surv(TGSZcut$AMLT,TGSZcut$AMLTC)[which
(TGSZcut$MF==0&TGSZcut$AMLTC>=0)],colSums(TGSZcut_pred[i*365/25,3:4,which(TGSZcut$MF==0&TGSZcut$AMLTC>=0)]),seq(100,365*25,50
))$iauc, "\n")
}

for(i in c(5,10,15,20)){
cat("Uno. EFS in CP at",i,"yrs",
AUC.uno(Surv(TGSZcut$EFS,TGSZcut$EFSC)[which(TGSZcut$MF==0&TGSZcut$EFSC>=0)],Surv(TGSZcut$EFS,TGSZcut$EFSC)[which(TGSZcut$
MF==0&TGSZcut$EFSC>=0)],colSums(TGSZcut_pred[i*365/25,1:5,which(TGSZcut$MF==0&TGSZcut$EFSC>=0)]),seq(100,365*25,50))$iauc, "\n")
}

for(i in c(5,10,15,20)){
cat("Uno. OS in ET at",i,"yrs",
AUC.uno(Surv(TGSZcut$Death,TGSZcut$DeathC)[which(TGSZcut$ET==1&TGSZcut$DeathC>=0)],Surv(TGSZcut$Death,TGSZcut$DeathC)[which(TG
SZcut$ET==1&TGSZcut$DeathC>=0)],colSums(TGSZcut_pred[i*365/25,1:4,which(TGSZcut$ET==1&TGSZcut$DeathC>=0)]),seq(100,365*25,50))$iauc
, "\n")
```

```
}

for(i in c(5,10,15,20)){
cat("Uno. MFT in ET at",i,"yrs",
AUC.uno(Surv(TGSZcut$MFT,TGSZcut$MFTC)[which(TGSZcut$ET==1&TGSZcut$MFTC>=0)],Surv(TGSZcut$MFT,TGSZcut$MFTC)[which(TGSZc
ut$ET==1&TGSZcut$MFTC>=0)],colSums(TGSZcut_pred[i*365/25,c(2,4,5),which(TGSZcut$ET==1&TGSZcut$MFTC>=0)]),seq(100,365*25,50))$iauc,
"\n")
}

for(i in c(5,10,15,20)){
cat("Uno. AML in ET at",i,"yrs",
AUC.uno(Surv(TGSZcut$AMLT,TGSZcut$AMLTC)[which(TGSZcut$ET==1&TGSZcut$AMLTC>=0)],Surv(TGSZcut$AMLT,TGSZcut$AMLTC)[which
(TGSZcut$ET==1&TGSZcut$AMLTC>=0)],colSums(TGSZcut_pred[i*365/25,3:4,which(TGSZcut$ET==1&TGSZcut$AMLTC>=0)]),seq(100,365*25,50)
)$iauc, "\n")
}

for(i in c(5,10,15,20)){
cat("Uno. EFS in ET at",i,"yrs",
AUC.uno(Surv(TGSZcut$EFS,TGSZcut$EFSC)[which(TGSZcut$ET==1&TGSZcut$EFSC>=0)],Surv(TGSZcut$EFS,TGSZcut$EFSC)[which(TGSZcut$E
T==1&TGSZcut$EFSC>=0)],colSums(TGSZcut_pred[i*365/25,1:5,which(TGSZcut$ET==1&TGSZcut$EFSC>=0)]),seq(100,365*25,50))$iauc, "\n")
}


for(i in c(5,10,15,20)){
cat("Uno. EFS in MF at",i,"yrs",
AUC.uno(Surv(TGSZcut$Death,TGSZcut$DeathC)[which(TGSZcut$MF==1&TGSZcut$DeathC>=0)],Surv(TGSZcut$Death,TGSZcut$DeathC)[which(TG
SZcut$MF==1&TGSZcut$DeathC>=0)],colSums(TGSZcut_pred[i*365/25,6:7,which(TGSZcut$MF==1&TGSZcut$DeathC>=0)]),seq(100,365*25,50))$iau
c, "\n")
}
```

```
for(i in c(5,10,15,20)){
cat("Uno. AML in MF at",i,"yrs",
AUC.uno(Surv(TGSZcut$AMLT,TGSZcut$AMLTC)[which(TGSZcut$MF==1&TGSZcut$AMLTC>=0)],Surv(TGSZcut$AMLT,TGSZcut$AMLTC)[which
(TGSZcut$MF==1&TGSZcut$AMLTC>=0)],TGSZcut_pred[i*365/25,7,which(TGSZcut$MF==1&TGSZcut$AMLTC>=0)],seq(100,365*25,50))$iauc,
"\n")
}


for(i in c(5,10,15,20)){
cat("Uno. OS in external CP at",i,"yrs",
AUC.uno(Surv(florCP$Death,florCP$DeathC)[which(florCP$MF==0&florCP$DeathC>=0)],Surv(florCP$Death,florCP$DeathC)[which(florCP$MF==0&fl
orCP$DeathC>=0)],colSums(florCP_pred[i*365/25,1:4,which(florCP$MF==0&florCP$DeathC>=0)]),seq(100,365*25,50))$iauc, "\n")
}

for(i in c(5,10,15,20)){
cat("Uno. MFT in external CP at",i,"yrs",
AUC.uno(Surv(florCP$MFT,florCP$MFTC)[which(florCP$MF==0&florCP$MFTC>=0)],Surv(florCP$MFT,florCP$MFTC)[which(florCP$MF==0&florC
P$MFTC>=0)],colSums(florCP_pred[i*365/25,c(2,4,5),which(florCP$MF==0&florCP$MFTC>=0)]),seq(100,365*25,50))$iauc, "\n")
}

for(i in c(5,10,15,20)){
cat("Uno. AML in external CP at",i,"yrs",
AUC.uno(Surv(florCP$AMLT,florCP$AMLTC)[which(florCP$MF==0&florCP$AMLTC>=0)],Surv(florCP$AMLT,florCP$AMLTC)[which(florCP$MF==
0&florCP$AMLTC>=0)],colSums(florCP_pred[i*365/25,3:4,which(florCP$MF==0&florCP$AMLTC>=0)]),seq(100,365*25,50))$iauc, "\n")
}

for(i in c(5,10,15,20)){
cat("Uno. EFS in external CP at",i,"yrs",
AUC.uno(Surv(florCP$EFS,florCP$EFSC)[which(florCP$MF==0&florCP$EFSC>=0)],Surv(florCP$EFS,florCP$EFSC)[which(florCP$MF==0&florCP$E
FSC>=0)],colSums(florCP_pred[i*365/25,1:5,which(florCP$MF==0&florCP$EFSC>=0)]),seq(100,365*25,50))$iauc, "\n")
```

```
}

for(i in c(5,10,15,20)){
cat("Uno. AML in MF at",i,"yrs",
AUC.uno(Surv(florMF$AMLT,florMF$AMLTC)[which(florMF$MF==1&florMF$AMLTC>=0)],Surv(florMF$AMLT,florMF$AMLTC)[which(florMF$M
F==1&florMF$AMLTC>=0)],florMF_pred[i*365/25,7,which(florMF$MF==1&florMF$AMLTC>=0)],seq(100,365*25,50))$iauc, "\n")
}

for(i in c(5,10,15,20)){
cat("Uno. EFS in MF at",i,"yrs",
AUC.uno(Surv(florMF$Death,florMF$DeathC)[which(florMF$MF==1&florMF$DeathC>=0)],Surv(florMF$Death,florMF$DeathC)[which(florMF$MF==
1&florMF$DeathC>=0)],colSums(florMF_pred[i*365/25,6:7,which(florMF$MF==1&florMF$DeathC>=0)]),seq(100,365*25,50))$iauc, "\n")
}

sink()



#Sediments plots
#Restrict to only patients who had event or sufficiently long FU to assess model
#Best and worst

TGSZCP<-TGSZcut[which(TGSZcut$MF==0&(TGSZcut$EFSC==1|TGSZcut$Death>=3650)),]
TGSZCPpred<-TGSZcut_pred[,,which(TGSZcut$MF==0&(TGSZcut$EFSC==1|TGSZcut$Death>=3650))]
TGSZMF<-TGSZcut[which(TGSZcut$MF==1&(TGSZcut$EFSC==1|TGSZcut$Death>=5*365)),]
TGSZMFpred<-TGSZcut_pred[,,which(TGSZcut$MF==1&(TGSZcut$EFSC==1|TGSZcut$Death>=5*365))]

pdf("CP_OS_top30.pdf",width=2)
par(mfrow=c(10,3), mar=c(0,0,0,0))
for(i in order(colSums(TGSZCPpred[10*365/25,1:4,]))[(nrow(TGSZCP)-29):nrow(TGSZCP)]){newdataplotLOO(TGSZCP,TGSZCPpred,i)}
```

```
dev.off()
pdf("CP_AML_top30.pdf",width=2)
par(mfrow=c(10,3), mar=c(0,0,0,0))
for(i in order(colSums(TGSZCPpred[10*365/25,3:4,]))[(nrow(TGSZCP)-29):nrow(TGSZCP)]){newdataplotLOO(TGSZCP,TGSZCPpred,i)}
dev.off()
pdf("CP_MF_top30.pdf",width=2)
par(mfrow=c(10,3), mar=c(0,0,0,0))
for(i in order(colSums(TGSZCPpred[10*365/25,c(2,4,5),]))[(nrow(TGSZCP)-29):nrow(TGSZCP)]){newdataplotLOO(TGSZCP,TGSZCPpred,i)}
dev.off()
pdf("CP_EFS_top30.pdf",width=2)
par(mfrow=c(10,3), mar=c(0,0,0,0))
for(i in order(colSums(TGSZCPpred[10*365/25,1:5,]))[(nrow(TGSZCP)-29):nrow(TGSZCP)]){newdataplotLOO(TGSZCP,TGSZCPpred,i)}
dev.off()
pdf("CP_EFS_best30.pdf",width=2)
par(mfrow=c(10,3), mar=c(0,0,0,0))
for(i in order(colSums(TGSZCPpred[20*365/25,1:5,]))[1:30]){newdataplotLOO(TGSZCP,TGSZCPpred,i)}
dev.off()
pdf("CP_transdeaths_top30.pdf",width=2)
par(mfrow=c(10,3), mar=c(0,0,0,0))
for(i in order(colSums(TGSZCPpred[10*365/25,2:4,]))[(nrow(TGSZCP)-29):nrow(TGSZCP)]){newdataplotLOO(TGSZCP,TGSZCPpred,i)}
dev.off()
pdf("MF_OS_top30.pdf",width=2)
par(mfrow=c(10,3), mar=c(0,0,0,0))
for(i in order(colSums(TGSZMFpred[10*365/25,6:7,]))[(nrow(TGSZMF)-29):nrow(TGSZMF)]){newdataplotLOO(TGSZMF,TGSZMFpred,i)}
dev.off()
pdf("MF_AML_top30.pdf",width=2)
par(mfrow=c(10,3), mar=c(0,0,0,0))
for(i in order(TGSZMFpred[10*365/25,7,])[(nrow(TGSZMF)-29):nrow(TGSZMF)]){newdataplotLOO(TGSZMF,TGSZMFpred,i)}
dev.off()
pdf("MF_OS_best30.pdf",width=2)
```

```
par(mfrow=c(10,3), mar=c(0,0,0,0))
for(i in order(colSums(TGSZMFpred[10*365/25,6:7,]))[1:30]){newdataplotLOO(TGSZMF,TGSZMFpred,i)}
dev.off()

# Flor MF sediment plots by HMR

  0   1   2   3
 43  89 109  57

pdf("DIPSS0.pdf",width=5)
par(mfrow=c(9,5), mar=c(0,0,0,0))
for(i in
order(colSums(florMF_pred[10*365/25,6:7,which(florMF$DIPSS==0)]))){newdataplotLOO(florMF[which(florMF$DIPSS==0),],florMF_pred[,,which(flor
MF$DIPSS==0)],i)}
dev.off()

pdf("DIPSS1.pdf",width=10)
par(mfrow=c(9,10), mar=c(0,0,0,0))
for(i in
order(colSums(florMF_pred[10*365/25,6:7,which(florMF$DIPSS==1)]))){newdataplotLOO(florMF[which(florMF$DIPSS==1),],florMF_pred[,,which(flor
MF$DIPSS==1)],i)}
dev.off()

pdf("DIPSS2.pdf",width=12)
par(mfrow=c(9,12), mar=c(0,0,0,0))
for(i in
order(colSums(florMF_pred[10*365/25,6:7,which(florMF$DIPSS==2)]))){newdataplotLOO(florMF[which(florMF$DIPSS==2),],florMF_pred[,,which(flor
MF$DIPSS==2)],i)}
dev.off()
```

```
pdf("DIPSS3.pdf",width=7)
par(mfrow=c(9,7), mar=c(0,0,0,0))
for(i in
order(colSums(florMF_pred[10*365/25,6:7,which(florMF$DIPSS==3)]))){newdataplotLOO(florMF[which(florMF$DIPSS==3),],florMF_pred[,,which(flor
MF$DIPSS==3)],i)}
dev.off()

#Sediment plots
# 30 equally spread

pdf("CP_EFS_range.pdf",width=3)
par(mfrow=c(10,4), mar=c(0,0,0,0))
for(i in order(colSums(TGSZCPpred[20*365/25,1:5,]))[seq(1,857,length.out=50)]){newdataplotLOO(TGSZCP,TGSZCPpred,i)}
dev.off()

pdf("MF_EFS_range40.pdf",width=3)
par(mfrow=c(10,4), mar=c(0,0,0,0))
for(i in order(colSums(TGSZMFpred[10*365/25,6:7,]))[seq(1,210,length.out=40)]){newdataplotLOO(TGSZMF,TGSZMFpred,i)}
dev.off()
```

Appendix 3: Mutational profiles of 2035 MPN patients

Patient	MPN	Gene	Chr	Pos	WT	MT	Protein	VAF	Mutation	JAK2/CALR/MPL status	CN aberration
1	ET	TET2	4	106155798	T	G	p.Y233*	0.09	Nonsense	JAK2V617F,MPLW515	-
1	ET	SRSF2	17	74732959	G	T	p.P95H	0.07	Missense	JAK2V617F,MPLW515	-
1	ET	ASXL1	20	31023702	C	T	p.Q1063*	0.38	Nonsense	JAK2V617F,MPLW515	-
2	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
3	ET	-	-	-	-	-	-	-	-	CALR	-
4	ET	DNMT3A	2	25464460	C	T	p.G685R	0.38	Missense	JAK2V617F	-
5	ET	-	-	-	-	-	-	-	-	-	-
6	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
7	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
8	ET	NFE2	12	54686554	_	delA	p.I242fs*6	0.45	frameshift (+1bp)	MPLW515	-
9	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
10	ET	DNMT3A	2	25457242	C	T	p.R882H	0.15	Missense	JAK2V617F	-
10	ET	IDH2	15	90631934	C	T	p.R140Q	0.27	Missense	JAK2V617F	-
11	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
12	ET	-	-	-	-	-	-	-	-	CALR	-
13	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
14	ET	-	-	-	-	-	-	-	-	-	-
15	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
16	ET	-	-	-	-	-	-	-	-	JAK2V617F	Chr9
17	ET	-	-	-	-	-	-	-	-	CALR	-
18	ET	-	-	-	-	-	-	-	-	CALR	-

19	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
20	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
21	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
22	ET	-	-	-	-	-	-	-	-	-	-
23	ET	-	-	-	-	-	-	-	-	CALR	Chr19
24	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
25	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
26	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
27	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
28	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
29	ET	-	-	-	-	-	-	-	-	CALR	-
30	ET	-	-	-	-	-	-	-	-	CALR	-
31	ET	-	-	-	-	-	-	-	-	CALR	-
32	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
33	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
34	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
35	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
36	ET	TET2	4	106155921	_	delC	p.N275fs*18	0.02	frameshift (+1bp)	JAK2V617F	-
37	ET	-	-	-	-	-	-	-	-	JAK2V617F	Chr9
38	ET	-	-	-	-	-	-	-	-	-	-
39	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
40	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
41	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
42	ET	TET2	4	106190798	G	T	p.R1359L	0.47	Missense	JAK2V617F	-
43	ET	-	-	-	-	-	-	-	-	CALR	-
44	ET	-	-	-	-	-	-	-	-	JAK2V617F	-

45	ET	-	-	-	-	-	-	-	-	CALR	-
46	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
47	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
48	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
49	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
50	ET	-	-	-	-	-	-	-	-	CALR	-
51	ET	-	-	-	-	-	-	-	-	CALR	-
52	ET	-	-	-	-	-	-	-	-	-	-
53	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
54	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
55	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
56	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
57	ET	TET2	4	106162499	_	delA	p.I1139fs*13	0.16	frameshift (+1bp)	JAK2V617F	Chr9
57	ET	TET2	4	106164023	_	insA	p.R1179fs*7	0.17	frameshift (+2bp)	JAK2V617F	Chr9
58	ET	DNMT3A	2	25470560	C	T	p.W305*	0.2	Nonsense	JAK2V617F	-
58	ET	GNAS	20	57484420	C	A	p.R201S	0.21	Missense	JAK2V617F	-
59	ET	U2AF1	21	44514777	T	C	p.Q157R	0.18	Missense	-	-
60	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
61	ET	-	-	-	-	-	-	-	-	-	-
62	ET	TET2	4	106157212	C	T	p.Q705*	0.08	Nonsense	JAK2V617F	-
63	ET	-	-	-	-	-	-	-	-	JAK2V617F	Chr9
64	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
65	ET	-	-	-	-	-	-	-	-	CALR	-
66	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
67	ET	-	-	-	-	-	-	-	-	CALR	-
68	ET	-	-	-	-	-	-	-	-	MPLW515	-

69	ET	-	-	-	-	-	-	-	-	CALR	-
70	ET	-	-	-	-	-	-	-	-	JAK2V617F	Chr9
71	ET	DNMT3A	2	25457242	C	T	p.R882H	0.14	Missense	MPLW515	Chr1p
72	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
73	ET	TET2	4	106197365	_	delG	p.V1900fs*8	0.04	frameshift (+1bp)	JAK2V617F	-
74	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
75	ET	DNMT3A	2	25459796	G	A	p.?	0.56	Splice	CALR	-
76	ET	DNMT3A	2	25463552	_	delG	p.C710fs*1	0.53	frameshift (+1bp)	JAK2V617F	Chr20
76	ET	U2AF1	21	44524456	G	T	p.S34Y	0.49	Missense	JAK2V617F	Chr20
77	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
78	ET	-	-	-	-	-	-	-	-	-	-
79	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
80	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
81	ET	-	-	-	-	-	-	-	-	CALR	Chr20
82	ET	TET2	4	106197269	C	T	p.H1868Y	0.49	Missense	CALR	Chr20
83	ET	-	-	-	-	-	-	-	-	CALR	-
84	ET	ASXL1	20	31022675	_	delC	p.L721fs*4	0.33	frameshift (+1bp)	JAK2V617F	-
85	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
86	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
87	ET	DNMT3A	2	25470498	G	A	p.R326C	0.51	Missense	JAK2V617F,MPLS505N	-
87	ET	TET2	4	106157240	C	A	p.S714*	0.45	Nonsense	JAK2V617F,MPLS505N	-
88	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
89	ET	U2AF1	21	44514777	T	G	p.Q157P	0.48	Missense	JAK2V617F	-
90	ET	-	-	-	-	-	-	-	-	CALR	-
91	ET	-	-	-	-	-	-	-	-	MPLW515	-
92	ET	-	-	-	-	-	-	-	-	JAK2V617F	-

93	ET	-	-	-	-	-	-	-	-	CALR	-
94	ET	-	-	-	-	-	-	-	-	JAK2V617F	Chr9
95	ET	SF3B1	2	198267359	C	A	p.K666N	0.33	Missense	JAK2V617F	Chr9
95	ET	TET2	4	106197317	A	G	p.T1884A	0.36	Missense	JAK2V617F	Chr9
96	ET	-	-	-	-	-	-	-	-	-	-
97	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
98	ET	-	-	-	-	-	-	-	-	-	-
99	ET	DNMT3A	2	25457242	C	T	p.R882H	0.42	Missense	JAK2V617F	-
100	ET	EZH2	7	148544318	G	A	p.R25*	0.77	Nonsense	JAK2V617F	Chr7,9
100	ET	ASXL1	20	31021535	C	T	p.Q512*	0.46	Nonsense	JAK2V617F	Chr7,9
101	ET	-	-	-	-	-	-	-	-	JAK2V617F	Chr9
102	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
103	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
104	ET	-	-	-	-	-	-	-	-	-	-
105	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
106	ET	DNMT3A	2	25457242	C	T	p.R882H	0.17	Missense	JAK2V617F	-
107	ET	-	-	-	-	-	-	-	-	CALR	-
108	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
109	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
110	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
111	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
112	ET	PPM1D	17	58740479	C	T	p.Q462*	0.15	Nonsense	-	-
113	ET	-	-	-	-	-	-	-	-	-	-
114	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
115	ET	TET2	4	106164793	T	A	p.C1221S	0.38	Missense	JAK2V617F	-
116	ET	-	-	-	-	-	-	-	-	CALR	-

117	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
118	ET	-	-	-	-	-	-	-	-	JAK2V617F	Chr11
119	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
120	ET	-	-	-	-	-	-	-	-	-	-
121	ET	TET2	4	106157755	C	T	p.Q886*	0.18	Nonsense	JAK2V617F	-
122	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
123	ET	NFE2	12	54686495	_	delCTCT	p.E261fs*3	0.12	frameshift (+1bp)	CALR	Chr20
124	ET	-	-	-	-	-	-	-	-	MPLW515	-
125	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
126	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
127	ET	-	-	-	-	-	-	-	-	-	-
128	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
129	ET	TP53	17	7577121	G	C	p.R273G	0.42	Missense	JAK2V617F	-
129	ET	ASXL1	20	31022284	_	insT	p.Y591fs*28	0.04	frameshift (+2bp)	JAK2V617F	-
130	ET	-	-	-	-	-	-	-	-	CALR	-
131	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
132	ET	DNMT3A	2	25466790	_	delGACAGCACCCG	p.R635fs*3	0.27	frameshift (+2bp)	JAK2V617F	-
134	ET	-	-	-	-	-	-	-	-	MPLW515	-
135	ET	-	-	-	-	-	-	-	-	JAK2V617F	Chr9
136	ET	DNMT3A	2	25469624	_	delTCCCC<13>CCACC	p.?	0.17	Splice	JAK2V617F	Chr11
136	ET	CBL	11	119148981	T	A	p.C401S	0.05	Missense	JAK2V617F	Chr11
137	ET	DNMT3A	2	25463228	_	delAAAGAGCCAG	p.F752fs*24	0.37	frameshift (+1bp)	JAK2V617F,CALR	-
138	ET	TET2	4	106196430	_	delT	p.I1588fs*8	0.1	frameshift (+1bp)	JAK2V617F	-
139	ET	TET2	4	106156690	C	T	p.Q531*	0.49	Nonsense	CALR	-
140	ET	-	-	-	-	-	-	-	-	-	-
141	ET	-	-	-	-	-	-	-	-	MPLW515	Chr1p

142	ET	PHF6	X	133511707	_	delT	p.C20fs*1	0.22	frameshift (+1bp)	CALR	-
143	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
144	ET	EZH2	7	148507424	C	G	p.?	0.18	Splice	CALR	-
145	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
146	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
147	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
148	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
149	ET	TP53	17	7578369	_	delA	p.?	0.31	Splice	JAK2V617F	-
150	ET	TET2	4	106156963	C	T	p.Q622*	0.41	Nonsense	JAK2V617F	-
151	ET	-	-	-	-	-	-	-	-	CALR	-
152	ET	-	-	-	-	-	-	-	-	CALR	-
153	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
154	ET	-	-	-	-	-	-	-	-	CALR	-
155	ET	DNMT3A	2	25457242	C	T	p.R882H	0.47	Missense	CALR	Chr20
156	ET	-	-	-	-	-	-	-	-	JAK2V617F,CALR	-
157	ET	PPM1D	17	58740510	_	delA	p.E472fs*11	0.02	frameshift (+1bp)	JAK2V617F	-
157	ET	PPM1D	17	58740668	G	T	p.E525*	0.05	Nonsense	JAK2V617F	-
158	ET	-	-	-	-	-	-	-	-	CALR	-
159	ET	-	-	-	-	-	-	-	-	JAK2V617F	Chr9
160	ET	-	-	-	-	-	-	-	-	CALR	-
161	ET	KRAS	12	25380283	C	A	p.A59S	0.1	Missense	JAK2V617F	Chr9
161	ET	ASXL1	20	31022937	_	delC	p.P808fs*10	0.03	frameshift (+1bp)	JAK2V617F	Chr9
162	ET	TET2	4	106197355	G	T	p.R1896S	0.89	Missense	JAK2V617F	Chr4,9
163	ET	TET2	4	106196213	C	T	p.R1516*	0.06	Nonsense	CALR	-
164	ET	-	-	-	-	-	-	-	-	JAK2V617F	Chr9,14
165	ET	TET2	4	106197036	_	delT	p.S1791fs*29	0.4	frameshift (+1bp)	CALR	-

165	ET	ASXL1	20	31023092	_	delC	p.N859fs*8	0.44	frameshift (+1bp)	CALR	-
166	ET	-	-	-	-	-	-	-	-	CALR	-
167	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
168	ET	-	-	-	-	-	-	-	-	CALR	-
169	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
170	ET	-	-	-	-	-	-	-	-	JAK2V617F	Chr9
171	ET	KRAS	12	25380256	T	A	p.R68W	0.07	Missense	JAK2V617F	-
172	ET	TET2	4	106190905	G	A	p.?	0.14	Splice	CALR	-
172	ET	TP53	17	7578388	C	T	p.R181H	0.23	Missense	CALR	-
172	ET	ASXL1	20	31022844	G	T	p.E777*	0.19	Nonsense	CALR	-
173	ET	NF1	17	29654554	G	A	p.R1769Q	0.56	Missense	CALR	-
173	ET	PPM1D	17	58740559	_	delT	p.S489fs*2	0.05	frameshift (+1bp)	CALR	-
173	ET	PPM1D	17	58740376	G	A	p.W427*	0.4	Nonsense	CALR	-
174	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
175	ET	DNMT3A	2	25467083	G	A	p.R598*	0.34	Nonsense	JAK2V617F	-
175	ET	TET2	4	106164020	T	G	p.I1177S	0.36	Missense	JAK2V617F	-
176	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
177	ET	TET2	4	106156439	_	delG	p.E448fs*2	0.45	frameshift (+1bp)	JAK2V617F	-
177	ET	TET2	4	106193931	C	T	p.R1465*	0.11	Nonsense	JAK2V617F	-
177	ET	ASXL1	20	31022937	_	delCC	p.P808fs*13	0.17	frameshift (+2bp)	JAK2V617F	-
178	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
179	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
180	ET	DNMT3A	2	25457242	C	T	p.R882H	0.34	Missense	JAK2V617F	-
180	ET	TET2	4	106157326	_	insA	p.Q744fs*10	0.11	frameshift (+2bp)	JAK2V617F	-
181	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
182	ET	-	-	-	-	-	-	-	-	JAK2V617F	-

183	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
184	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
185	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
186	ET	SF3B1	2	198266518	A	C	p.L773R	0.14	Missense	JAK2V617F	-
187	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
188	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
189	ET	DNMT3A	2	25457242	C	T	p.R882H	0.32	Missense	JAK2V617F	-
190	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
191	ET	-	-	-	-	-	-	-	-	CALR	-
192	ET	TET2	4	106196762	C	T	p.Q1699*	0.49	Nonsense	JAK2V617F	Chr9,14
192	ET	CBL	11	119149239	G	C	p.C416S	0.06	Missense	JAK2V617F	Chr9,14
192	ET	ZRSR2	X	15833805	C	A	p.S188*	0.92	Nonsense	JAK2V617F	Chr9,14
193	ET	NFE2	12	54686441	G	T	p.A280D	0.16	Missense	-	-
194	ET	-	-	-	-	-	-	-	-	-	-
195	ET	PPM1D	17	58740536	_	delG	p.A481fs*2	0.08	frameshift (+1bp)	CALR	Chr5,17,20
195	ET	PPM1D	17	58740542	_	insAC	p.R485fs*1	0.06	frameshift (+1bp)	CALR	Chr5,17,20
195	ET	TP53	17	7577120	C	T	p.R273H	0.5	Missense	CALR	Chr5,17,20
196	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
197	ET	ASXL1	20	31022937	_	delC	p.P808fs*10	0.03	frameshift (+1bp)	CALR	Chr20
198	ET	DNMT3A	2	25457242	C	T	p.R882H	0.31	Missense	JAK2V617F	-
198	ET	SF3B1	2	198267359	C	A	p.K666N	0.4	Missense	JAK2V617F	-
198	ET	GATA2	3	128200729	A	G	p.L359S	0.2	Missense	JAK2V617F	-
199	ET	PPM1D	17	58740508	_	delA	p.E472fs*11	0.01	frameshift (+1bp)	CALR	-
199	ET	PPM1D	17	58740546	T	G	p.L484*	0.16	Nonsense	CALR	-
199	ET	PPM1D	17	58740691	_	delA	p.N533fs*6	0.05	frameshift (+1bp)	CALR	-
200	ET	-	-	-	-	-	-	-	-	JAK2V617F	-

201	ET	ASXL1	20	31022899	_	delC	p.W796fs*22	0.36	frameshift (+1bp)	JAK2V617F	Chr12
202	ET	TET2	4	106164079	A	G	p.K1197E	0.12	Missense	JAK2V617F	Chr9
202	ET	CBL	11	119149245	T	C	p.F418S	0.06	Missense	JAK2V617F	Chr9
203	ET	SF3B1	2	198267359	C	A	p.K666N	0.43	Missense	JAK2V617F	Chr9
204	ET	-	-	-	-	-	-	-	-	CALR	-
205	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
206	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
207	ET	EZH2	7	148506401	C	T	p.?	0.07	Splice	JAK2V617F	-
208	ET	DNMT3A	2	25457242	C	T	p.R882H	0.33	Missense	JAK2V617F	-
209	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
210	ET	-	-	-	-	-	-	-	-	JAK2V617F,CALR	-
211	ET	-	-	-	-	-	-	-	-	CALR	-
212	ET	-	-	-	-	-	-	-	-	-	-
213	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
214	ET	TP53	17	7578503	C	T	p.V143M	0.05	Missense	CALR	-
215	ET	-	-	-	-	-	-	-	-	CALR	-
216	ET	TET2	4	106182915	_	delG	p.?	0.16	Splice	CALR	-
216	ET	TET2	4	106156359	_	delA	p.E421fs*6	0.03	frameshift (+1bp)	CALR	-
216	ET	NFE2	12	54687087	_	delA	p.Y65fs*46	0.18	frameshift (+1bp)	CALR	-
217	ET	-	-	-	-	-	-	-	-	CALR	-
218	ET	-	-	-	-	-	-	-	-	-	-
219	ET	TET2	4	106158055	G	T	p.G986*	0.05	Nonsense	CALR	-
220	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
221	ET	DNMT3A	2	25466836	_	delA	p.Y623fs*28	0.08	frameshift (+1bp)	CALR	-
222	ET	DNMT3A	2	25457242	C	T	p.R882H	0.24	Missense	CALR	-
223	ET	-	-	-	-	-	-	-	-	JAK2V617F	-

224	ET	-	-	-	-	-	-	-	-	CALR	-
225	ET	-	-	-	-	-	-	-	-	CALR	-
226	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
227	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
228	ET	-	-	-	-	-	-	-	-	CALR	Chr14
229	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
230	ET	-	-	-	-	-	-	-	-	CALR	-
231	ET	PPM1D	17	58740836	C	T	p.R581*	0.18	Nonsense	CALR	-
232	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
233	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
234	ET	-	-	-	-	-	-	-	-	CALR	-
235	ET	DNMT3A	2	25457243	G	A	p.R882C	0.39	Missense	CALR	-
235	ET	TET2	4	106156540	C	T	p.Q481*	0.07	Nonsense	CALR	-
236	ET	TET2	4	106196650	T	G	p.Y1661*	0.32	Nonsense	JAK2V617F	Chr9
237	ET	U2AF1	21	44524456	G	T	p.S34Y	0.14	Missense	MPLW515	-
238	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
239	ET	-	-	-	-	-	-	-	-	-	-
240	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
241	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
242	ET	-	-	-	-	-	-	-	-	JAK2V617F,MPLW515	-
243	ET	CBL	11	119149239	G	C	p.C416S	0.05	Missense	JAK2V617F	-
244	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
245	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
246	ET	MPL	1	43818306	T	G	p.Y591D	0.35	Missense	MPLW515	-
247	ET	-	-	-	-	-	-	-	-	-	-
248	ET	PPM1D	17	58740528	_	delGC	p.A479fs*1	0.13	frameshift (+2bp)	CALR	-

249	ET	DNMT3A	2	25457242	C	T	p.R882H	0.35	Missense	JAK2V617F	-
249	ET	IDH2	15	90631934	C	T	p.R140Q	0.38	Missense	JAK2V617F	-
250	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
251	ET	-	-	-	-	-	-	-	-	-	-
252	ET	-	-	-	-	-	-	-	-	MPLW515	Chr1p
253	ET	-	-	-	-	-	-	-	-	CALR	Chr19
254	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
255	ET	NF1	17	29664383	C	T	p.?	0.5	Splice	JAK2V617F	-
256	ET	TET2	4	106156819	C	T	p.Q574*	0.3	Nonsense	JAK2V617F,CALR	-
256	ET	PPM1D	17	58740498	C	G	p.S468*	0.31	Nonsense	JAK2V617F,CALR	-
257	ET	-	-	-	-	-	-	-	-	-	-
258	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
259	ET	TET2	4	106183003	C	A	p.Q1348K	0.3	Missense	JAK2V617F	-
260	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
261	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
262	ET	-	-	-	-	-	-	-	-	-	-
263	ET	DNMT3A	2	25457243	G	A	p.R882C	0.25	Missense	JAK2V617F	-
264	ET	PPM1D	17	58740836	C	T	p.R581*	0.06	Nonsense	-	-
265	ET	NFE2	12	54686495	_	delCTCT	p.E261fs*3	0.23	frameshift (+1bp)	CALR	-
266	ET	PPM1D	17	58740507	_	delCA	p.P471fs*4	0.11	frameshift (+2bp)	CALR	-
267	ET	-	-	-	-	-	-	-	-	CALR	-
268	ET	SF3B1	2	198267360	T	G	p.K666T	0.38	Missense	JAK2V617F	-
268	ET	ASXL1	20	31022937	_	insC	p.A809fs*13	0.09	frameshift (+2bp)	JAK2V617F	-
268	ET	ASXL1	20	31023272	_	delACCATCTGTTG	p.P920fs*24	0.01	frameshift (+2bp)	JAK2V617F	-
269	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
270	ET	-	-	-	-	-	-	-	-	JAK2V617F	-

271	ET	SF3B1	2	198267489	T	C	p.Y623C	0.46	Missense	CALR	-
271	ET	KRAS	12	25398262	C	G	p.L19F	0.31	Missense	CALR	-
272	ET	-	-	-	-	-	-	-	-	-	-
273	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
274	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
275	ET	TET2	4	106180784	_	insG	p.C1271fs*29	0.07	frameshift (+2bp)	JAK2V617F	-
276	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
277	ET	-	-	-	-	-	-	-	-	CALR	-
278	ET	PPM1D	17	58740691	_	delA	p.N533fs*6	0.24	frameshift (+1bp)	-	-
279	ET	ASXL1	20	31022403	_	delCACCA<13>GCGGC	p.E635fs*15	0.34	frameshift (+2bp)	CALR	-
280	ET	-	-	-	-	-	-	-	-	CALR	-
281	ET	GNAS	20	57415346	G	A	p.R62Q	0.39	Missense	CALR	-
282	ET	NFE2	12	54686495	_	delCTCT	p.E261fs*3	0.33	frameshift (+1bp)	MPLW515	Chr1p
283	ET	DNMT3A	2	25463556	G	C	p.P709R	0.05	Missense	JAK2V617F	-
284	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
285	ET	ASXL1	20	31022287	_	insA	p.Y591fs*1	0.06	frameshift (+2bp)	JAK2V617F	Chr9
285	ET	ZRSR2	X	15836766	G	A	p.?	0.45	Splice	JAK2V617F	Chr9
286	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
287	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
288	ET	-	-	-	-	-	-	-	-	CALR	-
289	ET	TET2	4	106193892	C	T	p.R1452*	0.05	Nonsense	CALR	-
290	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
291	ET	-	-	-	-	-	-	-	-	CALR	-
292	ET	DNMT3A	2	25470516	G	A	p.R320*	0.47	Nonsense	CALR	-
292	ET	TET2	4	106158339	_	delTT	p.L1081fs*22	0.22	frameshift (+2bp)	CALR	-
292	ET	KRAS	12	25378562	C	T	p.A146T	0.09	Missense	CALR	-

292	ET	NF1	17	29588850	T	G	p.S1567A	0.07	Missense	CALR	-
293	ET	-	-	-	-	-	-	-	-	-	-
294	ET	-	-	-	-	-	-	-	-	-	Chr20
295	ET	TET2	4	106180784	_	insG	p.C1271fs*29	0.11	frameshift (+2bp)	CALR	-
295	ET	-	-	-	-	-	-	-	-	CALR	-
296	ET	PPM1D	17	58740691	_	delA	p.N533fs*6	0.03	frameshift (+1bp)	MPLW515	Chr1p,14
297	ET	-	-	-	-	-	-	-	-	CALR	-
298	ET	TET2	4	106180784	_	insG	p.C1271fs*29	0.03	frameshift (+2bp)	JAK2V617F	-
298	ET	TET2	4	106157603	C	G	p.S835*	0.05	Nonsense	JAK2V617F	-
299	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
300	ET	-	-	-	-	-	-	-	-	-	-
301	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
302	ET	-	-	-	-	-	-	-	-	-	-
303	ET	TET2	4	106196347	T	G	p.Y1560*	0.24	Nonsense	CALR	-
304	ET	TET2	4	106197235	_	insT	p.P1857fs*2	0.32	frameshift (+2bp)	JAK2V617F,MPLW515	-
305	ET	MPL	1	43815011	C	G	p.Q516E	0.26	Missense	MPLW515	-
305	ET	TET2	4	106157020	_	delG	p.Q642fs*58	0.01	frameshift (+1bp)	MPLW515	-
306	ET	-	-	-	-	-	-	-	-	CALR	-
307	ET	IDH1	2	209113112	C	T	p.R132H	0.09	Missense	CALR	-
308	ET	-	-	-	-	-	-	-	-	-	-
309	ET	TET2	4	106156157	_	delG	p.C353fs*19	0.07	frameshift (+1bp)	JAK2V617F	-
309	ET	ASXL1	20	31022418	G	T	p.E635*	0.09	Nonsense	JAK2V617F	-
310	ET	-	-	-	-	-	-	-	-	CALR	-
311	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
312	ET	-	-	-	-	-	-	-	-	-	-
313	ET	-	-	-	-	-	-	-	-	JAK2V617F	-

314	ET	-	-	-	-	-	-	-	-	MPLW515	-
315	ET	TET2	4	106156701	_	delAAAC	p.N535fs*6	0.35	frameshift (+1bp)	JAK2V617F	-
316	ET	-	-	-	-	-	-	-	-	-	-
317	ET	-	-	-	-	-	-	-	-	CALR	-
318	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
319	ET	-	-	-	-	-	-	-	-	CALR	-
320	ET	MLL3	7	151882660	_	delC	p.E1689fs*28	0.07	frameshift (+1bp)	-	-
321	ET	-	-	-	-	-	-	-	-	CALR	-
322	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
323	ET	-	-	-	-	-	-	-	-	CALR	-
324	ET	TET2	4	106162499	_	delA	p.I1139fs*13	0.05	frameshift (+1bp)	CALR	-
325	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
326	ET	-	-	-	-	-	-	-	-	CALR	-
327	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
328	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
329	ET	PPM1D	17	58740357	C	G	p.S421*	0.18	Nonsense	JAK2V617F	-
329	ET	ASXL1	20	31023231	_	delG	p.E906fs*2	0.07	frameshift (+1bp)	JAK2V617F	-
330	ET	TET2	4	106197317	A	G	p.T1884A	0.27	Missense	JAK2V617F	-
331	ET	TET2	4	106197285	T	C	p.I1873T	0.36	Missense	JAK2V617F	-
331	ET	SRSF2	17	74732959	G	T	p.P95H	0.31	Missense	JAK2V617F	-
331	ET	RUNX1	21	36259156	A	T	p.L112Q	0.34	Missense	JAK2V617F	-
332	ET	-	-	-	-	-	-	-	-	-	-
333	ET	-	-	-	-	-	-	-	-	CALR	-
334	ET	TET2	4	106164778	C	T	p.R1216*	0.45	Nonsense	JAK2V617F	-
335	ET	NFE2	12	54686511	_	delA	p.Y257fs*3	0.23	frameshift (+1bp)	MPLW515	-
335	ET	IDH2	15	90631934	C	T	p.R140Q	0.26	Missense	MPLW515	-

336	ET	-	-	-	-	-	-	-	-	CALR	-
337	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
338	ET	-	-	-	-	-	-	-	-	CALR	-
339	ET	-	-	-	-	-	-	-	-	-	-
340	ET	-	-	-	-	-	-	-	-	CALR	-
341	ET	-	-	-	-	-	-	-	-	CALR	-
342	ET	-	-	-	-	-	-	-	-	CALR	Chr9
343	ET	TET2	4	106155427	A	T	p.K110*	0.14	Nonsense	-	-
344	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
345	ET	-	-	-	-	-	-	-	-	CALR	Chr19
346	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
347	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
348	ET	DNMT3A	2	25470484	C	T	p.W330*	0.36	Nonsense	CALR	-
349	ET	-	-	-	-	-	-	-	-	CALR	-
350	ET	TET2	4	106158041	_	insT	p.K982fs*1	0.39	frameshift (+2bp)	JAK2V617F	Chr9
350	ET	MLL3	7	151882660	_	delC	p.E1689fs*28	0.06	frameshift (+1bp)	JAK2V617F	Chr9
351	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
352	ET	-	-	-	-	-	-	-	-	CALR	-
353	ET	-	-	-	-	-	-	-	-	-	-
354	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
355	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
356	ET	GNB1	1	1747229	T	C	p.K57E	0.43	Missense	JAK2V617F	-
356	ET	SH2B3	12	111885289	C	T	p.Q393*	0.06	Nonsense	JAK2V617F	-
357	ET	SF3B1	2	198266834	T	C	p.K700E	0.44	Missense	JAK2V617F	Chr1q,9
357	ET	TET2	4	106158489	_	insC	p.S1132fs*10	0.07	frameshift (+2bp)	JAK2V617F	Chr1q,9
358	ET	GNB1	1	1747229	T	C	p.K57E	0.05	Missense	JAK2V617F	-

359	ET	-	-	-	-	-	-	-	-	CALR	-
360	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
361	ET	MPL	1	43814967	T	C	p.V501A	0.63	Missense	MPLS505N	Chr1p
362	ET	TET2	4	106157002	C	T	p.Q635*	0.46	Nonsense	CALR	Chr9,19
362	ET	NFE2	12	54686562	_	insC	p.D240fs*10	0.44	frameshift (+2bp)	CALR	Chr9,19
363	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
364	ET	TET2	4	106157345	_	delAAATA	p.I750fs*2	0.46	frameshift (+2bp)	JAK2V617F	Chr9
365	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
366	ET	-	-	-	-	-	-	-	-	CALR	-
367	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
368	ET	-	-	-	-	-	-	-	-	CALR	-
369	ET	-	-	-	-	-	-	-	-	CALR	-
370	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
371	ET	ASXL1	20	31021348	C	G	p.Y449*	0.31	Nonsense	-	-
372	ET	DNMT3A	2	25463247	C	T	p.R749H	0.39	Missense	JAK2V617F	-
373	ET	-	-	-	-	-	-	-	-	CALR	-
374	ET	PTPN11	12	112924336	G	A	p.V428M	0.14	Missense	CALR	-
375	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
376	ET	-	-	-	-	-	-	-	-	CALR	-
377	ET	TET2	4	106197317	A	G	p.T1884A	0.17	Missense	CALR	-
378	ET	TET2	4	106156336	_	delC	p.P413fs*14	0.29	frameshift (+1bp)	JAK2V617F	-
378	ET	ASXL1	20	31022403	_	delCACCA<13>GCGGC	p.E635fs*15	0.29	frameshift (+2bp)	JAK2V617F	-
379	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
380	ET	-	-	-	-	-	-	-	-	CALR	-
381	ET	-	-	-	-	-	-	-	-	CALR	-
382	ET	-	-	-	-	-	-	-	-	CALR	-

383	ET	SRSF2	17	74732959	G	T	p.P95H	0.1	Missense	-	-
384	ET	DNMT3A	2	25470497	C	T	p.R326H	0.34	Missense	JAK2V617F	-
384	ET	TET2	4	106197285	T	C	p.I1873T	0.32	Missense	JAK2V617F	-
384	ET	TET2	4	106196657	C	T	p.Q1664*	0.05	Nonsense	JAK2V617F	-
385	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
386	ET	TET2	4	106157761	C	T	p.Q888*	0.31	Nonsense	JAK2V617F	-
387	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
389	ET	-	-	-	_	-	-	-	-	JAK2V617F	-
390	ET	-	-	-	-	-	-	-	-	-	-
391	ET	-	-	-	-	-	-	-	-	CALR	-
392	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
393	ET	-	-	-	-	-	-	-	-	-	-
394	ET	-	-	-	-	-	-	-	-	-	-
395	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
396	ET	TET2	4	106196329	_	insA	p.E1555fs*23	0.06	frameshift (+2bp)	JAK2V617F	-
396	ET	TET2	4	106164778	C	T	p.R1216*	0.32	Nonsense	JAK2V617F	-
397	ET	-	-	-	-	-	-	-	-	CALR	-
398	ET	SH2B3	12	111885522	_	insG	p.V434fs*22	0.11	frameshift (+2bp)	JAK2V617F	-
399	ET	ASXL1	20	31024148	_	delCTCC	p.L1213fs*3	0.02	frameshift (+1bp)	CALR	-
400	ET	-	-	-	-	-	-	-	-	CALR,MPLW515	-
401	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
402	ET	-	-	-	-	-	-	-	-	CALR	-
403	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
404	ET	DNMT3A	2	25457242	C	T	p.R882H	0.41	Missense	JAK2V617F	-
404	ET	SF3B1	2	198266834	T	C	p.K700E	0.19	Missense	JAK2V617F	-
405	ET	ASXL1	20	31022403	_	delCACCA<13>GCGGC	p.E635fs*15	0.04	frameshift (+2bp)	JAK2V617F	-

406	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
407	ET	-	-	-	-	-	-	-	-	-	-
408	ET	-	-	-	-	-	-	-	-	-	-
409	ET	IDH2	15	90631934	C	T	p.R140Q	0.16	Missense	CALR	-
410	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
411	ET	SF3B1	2	198267359	C	A	p.K666N	0.34	Missense	-	-
412	ET	-	-	-	-	-	-	-	-	CALR	Chr20
413	ET	SF3B1	2	198267361	T	G	p.K666Q	0.38	Missense	JAK2V617F	-
414	ET	MLL3	7	151927007	C	G	p.?	0.06	Splice	CALR	-
415	ET	DNMT3A	2	25458595	A	G	p.W860R	0.23	Missense	JAK2V617F	-
416	ET	-	-	-	-	-	-	-	-	CALR	-
417	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
418	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
419	ET	PPM1D	17	58740526	_	delT	p.C478fs*5	0.17	frameshift (+1bp)	CALR	-
420	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
421	ET	-	-	-	-	-	-	-	-	-	-
422	ET	TP53	17	7578406	C	T	p.R175H	0.13	Missense	JAK2V617F	-
423	ET	TET2	4	106158218	T	A	p.L1040*	0.32	Nonsense	MPLW515	Chr1p
423	ET	TET2	4	106197355	G	T	p.R1896S	0.06	Missense	MPLW515	Chr1p
424	ET	-	-	-	-	-	-	-	-	-	-
425	ET	-	-	-	-	-	-	-	-	CALR	-
426	ET	TET2	4	106158478	C	T	p.Q1127*	0.1	Nonsense	JAK2V617F	-
426	ET	TET2	4	106156991	C	A	p.S631*	0.07	Nonsense	JAK2V617F	-
427	ET	-	-	-	-	-	-	-	-	-	-
428	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
429	ET	-	-	-	-	-	-	-	-	-	-

430	ET	MLL3	7	151970896	A	T	p.C302*	0.04	Nonsense	JAK2V617F,CALR	-
431	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
432	ET	PPM1D	17	58740653	C	T	p.Q520*	0.05	Nonsense	-	-
433	ET	MLL3	7	151962122	C	T	p.?	0.07	Splice	-	-
434	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
435	ET	RB1	13	49039230	C	T	p.Q770*	0.17	Nonsense	CALR	-
436	ET	-	-	-	-	-	-	-	-	CALR	-
437	ET	-	-	-	-	-	-	-	-	CALR	-
438	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
439	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
440	ET	DNMT3A	2	25458619	T	C	p.M852V	0.05	Missense	-	-
441	ET	ASXL1	20	31022238	C	T	p.Q575*	0.43	Nonsense	JAK2V617F	-
442	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
443	ET	-	-	-	-	-	-	-	-	-	-
444	ET	TET2	4	106193719	A	G	p.?	0.29	Splice	JAK2V617F	-
444	ET	ASXL1	20	31021730	_	insTGTT	p.?	0.35	Splice	JAK2V617F	-
445	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
446	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
447	ET	-	-	-	-	-	-	-	-	-	-
448	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
449	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
450	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
451	ET	-	-	-	-	-	-	-	-	CALR	-
452	ET	-	-	-	-	-	-	-	-	CALR	-
453	ET	DNMT3A	2	25463248	G	A	p.R749C	0.35	Missense	MPLW515	Chr1p
453	ET	TET2	4	106157371	C	T	p.Q758*	0.13	Nonsense	MPLW515	Chr1p

453	ET	EZH2	7	148511175	C	T	p.C576Y	0.09	Missense	MPLW515	Chr1p
453	ET	EZH2	7	148523590	C	T	p.R288Q	0.28	Missense	MPLW515	Chr1p
454	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
455	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
456	ET	DNMT3A	2	25463563	C	T	p.G707S	0.14	Missense	CALR	-
456	ET	TET2	4	106190858	C	T	p.A1379V	0.14	Missense	CALR	-
457	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
458	ET	-	-	-	-	-	-	-	-	CALR	-
459	ET	MBD1	18	47803325	G	A	p.S90L	0.05	Missense	-	-
460	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
461	ET	-	-	-	-	-	-	-	-	MPLW515	-
462	ET	-	-	-	-	-	-	-	-	CALR	Chr19
463	ET	-	-	-	-	-	-	-	-	-	-
464	ET	-	-	-	-	-	-	-	-	CALR	-
465	ET	DNMT3A	2	25463229	A	G	p.F755S	0.05	Missense	JAK2V617F	-
465	ET	SF3B1	2	198266834	T	C	p.K700E	0.05	Missense	JAK2V617F	-
466	ET	TET2	4	106193748	C	T	p.R1404*	0.06	Nonsense	JAK2V617F	-
467	ET	-	-	-	-	-	-	-	-	CALR	-
468	ET	SH2B3	12	111856361	_	insT	p.R139fs*33	0.06	frameshift (+2bp)	JAK2V617F	-
469	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
470	ET	-	-	-	-	-	-	-	-	-	-
471	ET	DNMT3A	2	25457242	C	T	p.R882H	0.18	Missense	JAK2V617F	-
471	ET	TP53	17	7577545	T	C	p.M246V	0.19	Missense	JAK2V617F	-
472	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
473	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
474	ET	-	-	-	-	-	-	-	-	JAK2V617F	-

475	ET	TET2	4	106158341	T	A	p.L1081*	0.29	Nonsense	JAK2V617F	-
476	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
477	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
478	ET	DNMT3A	2	25463298	_	delAAG	p.F732delF	0.37	Inframe	JAK2V617F	-
479	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
480	ET	-	-	-	-	-	-	-	-	-	-
481	ET	TET2	4	106158265	_	delG	p.V1056fs*10	0.23	frameshift (+1bp)	CALR	Chr20
482	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
483	ET	TET2	4	106157970	_	insA	p.Q958fs*14	0.25	frameshift (+2bp)	JAK2V617F	-
484	ET	-	-	-	-	-	-	-	-	CALR	-
485	ET	-	-	-	-	-	-	-	-	CALR	-
486	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
487	ET	-	-	-	-	-	-	-	-	MPLW515	-
488	ET	-	-	-	-	-	-	-	-	CALR	-
489	ET	EZH2	7	148544338	C	T	p.R18H	0.48	Missense	CALR	Chr14
490	ET	-	-	-	-	-	-	-	-	CALR	-
491	ET	-	-	-	-	-	-	-	-	-	-
492	ET	TP53	17	7577061	C	A	p.G293W	0.46	Missense	JAK2V617F	Chr17
493	ET	ASXL1	20	31023630	C	T	p.Q1039*	0.14	Nonsense	JAK2V617F	-
494	ET	-	-	-	-	-	-	-	-	CALR	-
495	ET	-	-	-	-	-	-	-	-	JAK2V617F	Chr9
496	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
497	ET	IDH2	15	90631934	C	T	p.R140Q	0.42	Missense	MPLS505N	-
497	ET	SRSF2	17	74732959	G	C	p.P95R	0.44	Missense	MPLS505N	-
498	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
499	ET	-	-	-	-	-	-	-	-	JAK2V617F	-

500	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
501	ET	DNMT3A	2	25462004	C	T	p.M801I	0.06	Missense	CALR	-
502	ET	DNMT3A	2	25468177	_	delAGGC	p.S499fs*151	0.23	frameshift (+1bp)	JAK2V617F	-
503	ET	DNMT3A	2	25464460	C	T	p.G685R	0.14	Missense	JAK2V617F	-
504	ET	-	-	-	-	-	-	-	-	-	-
505	ET	-	-	-	-	-	-	-	-	-	-
506	ET	-	-	-	-	-	-	-	-	JAK2V617F,CALR	-
507	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
508	ET	MPL	1	43818310	G	A	p.R592Q	0.11	Missense	JAK2V617F	-
509	ET	-	-	-	-	-	-	-	-	CALR	-
510	ET	-	-	-	-	-	-	-	-	CALR	-
511	ET	-	-	-	-	-	-	-	-	CALR	-
512	ET	-	-	-	-	-	-	-	-	CALR	-
513	ET	-	-	-	-	-	-	-	-	CALR	-
514	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
515	ET	-	-	-	-	-	-	-	-	CALR	-
516	ET	-	-	-	-	-	-	-	-	CALR	-
517	ET	-	-	-	-	-	-	-	-	CALR	-
518	ET	TET2	4	106196424	C	G	p.S1586*	0.17	Nonsense	JAK2V617F	-
519	ET	-	-	-	-	-	-	-	-	JAK2V617F	Chr9
520	ET	TET2	4	106196886	T	G	p.L1740*	0.13	Nonsense	JAK2V617F	-
521	ET	-	-	-	-	-	-	-	-	CALR	-
522	ET	TET2	4	106190894	G	T	p.G1391V	0.25	Missense	JAK2V617F	-
523	ET	-	-	-	-	-	-	-	-	CALR	-
524	ET	-	-	-	-	-	-	-	-	MPLS204P	Chr1p
525	ET	-	-	-	-	-	-	-	-	JAK2V617F	-

526	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
527	ET	TET2	4	106193723	_	insA	p.C1396fs*5	0.43	frameshift (+2bp)	JAK2V617F	-
528	ET	-	-	-	-	-	-	-	-	CALR	-
529	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
530	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
531	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
532	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
533	ET	-	-	-	-	-	-	-	-	CALR	-
534	ET	-	-	-	-	-	-	-	-	CALR	-
535	ET	-	-	-	-	-	-	-	-	CALR	-
536	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
537	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
538	ET	KIT	4	55594198	G	A	p.R634Q	0.56	Missense	JAK2V617F	-
539	ET	-	-	-	-	-	-	-	-	CALR	Chr20
540	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
541	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
542	ET	SF3B1	2	198267359	C	A	p.K666N	0.4	Missense	JAK2V617F	-
543	ET	PPM1D	17	58740522	_	delA	p.N477fs*6	0.1	frameshift (+1bp)	JAK2V617F	-
543	ET	PPM1D	17	58740642	C	A	p.S516*	0.19	Nonsense	JAK2V617F	-
544	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
545	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
546	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
547	ET	TP53	17	7578413	C	T	p.V173M	0.06	Missense	JAK2V617F	Chr17
548	ET	-	-	-	-	-	-	-	-	CALR	-
549	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
550	ET	-	-	-	-	-	-	-	-	-	-

551	ET	-	-	-	-	-	-	-	-	-	-
552	ET	-	-	-	-	-	-	-	-	-	-
553	ET	MLL3	7	151927096	_	delC	p.C963fs*68	0.09	frameshift (+1bp)	-	-
554	ET	-	-	-	-	-	-	-	-	JAK2V617F,CALR	-
555	ET	-	-	-	-	-	-	-	-	JAK2V617F	Chr9
556	ET	-	-	-	-	-	-	-	-	CALR	-
557	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
558	ET	TP53	17	7577509	C	T	p.E258K	0.37	Missense	CALR	Chr17
558	ET	U2AF1	21	44514777	T	C	p.Q157R	0.18	Missense	CALR	Chr17
559	ET	TET2	4	106197082	_	delC	p.H1806fs*14	0.14	frameshift (+1bp)	MPLW515	-
560	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
561	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
562	ET	DNMT3A	2	25457242	C	T	p.R882H	0.44	Missense	JAK2V617F	-
563	ET	-	-	-	-	-	-	-	-	MPLW515	-
564	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
565	ET	EZH2	7	148526906	T	C	p.Y133C	0.13	Missense	CALR	-
565	ET	CEBPA	19	33792303	C	T	p.G340S	0.51	Missense	CALR	-
566	ET	-	-	-	-	-	-	-	-	CALR	-
567	ET	-	-	-	-	-	-	-	-	CALR	-
568	ET	IDH2	15	90631934	C	A	p.R140L	0.38	Missense	CALR	Chr18
568	ET	SRSF2	17	74732959	G	A	p.P95L	0.35	Missense	CALR	Chr18
569	ET	-	-	-	-	-	-	-	-	CALR	-
570	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
571	ET	-	-	-	-	-	-	-	-	-	-
572	ET	-	-	-	-	-	-	-	-	CALR	-
573	ET	MLL3	7	151970872	A	T	p.C310*	0.05	Nonsense	-	-

574	ET	TET2	4	106190785	GC	AT	p.A1355I	0.05	Missense	JAK2V617F	-
574	ET	TET2	4	106193731	_	insT	p.T1399fs*2	0.04	frameshift (+2bp)	JAK2V617F	-
575	ET	-	-	-	-	-	-	-	-	CALR	-
576	ET	-	-	-	-	-	-	-	-	-	-
577	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
578	ET	PPM1D	17	58740665	C	T	p.Q524*	0.1	Nonsense	JAK2V617F	-
579	ET	-	-	-	-	-	-	-	-	CALR	-
580	ET	-	-	-	-	-	-	-	-	MPLW515	-
581	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
582	ET	-	-	-	-	-	-	-	-	JAK2V617F	Chr9
583	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
584	ET	-	-	-	-	-	-	-	-	CALR	-
585	ET	-	-	-	-	-	-	-	-	-	-
586	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
587	ET	-	-	-	-	-	-	-	-	JAK2V617F,MPLW515	-
588	ET	TET2	4	106155921	_	delC	p.N275fs*18	0.5	frameshift (+1bp)	CALR	-
589	ET	-	-	-	-	-	-	-	-	-	-
590	ET	-	-	-	-	-	-	-	-	JAK2V617F,CALR	-
591	ET	-	-	-	-	-	-	-	-	CALR	-
592	ET	-	-	-	-	-	-	-	-	MPLW515	-
593	ET	-	-	-	-	-	-	-	-	JAK2V617F	Chr17
594	ET	TET2	4	106196382	G	A	p.R1572Q	0.49	Missense	JAK2V617F	-
595	ET	-	-	-	-	-	-	-	-	CALR	Chr14
596	ET	-	-	-	-	-	-	-	-	-	-
597	ET	SF3B1	2	198267359	C	A	p.K666N	0.18	Missense	JAK2V617F	-
597	ET	TET2	4	106196631	_	delCGATGGAT	p.M1656fs*2	0.16	frameshift (+2bp)	JAK2V617F	-

598	ET	DNMT3A	2	25467083	G	A	p.R598*	0.1	Nonsense	CALR	-
599	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
600	ET	-	-	-	-	-	-	-	-	-	-
601	ET	DNMT3A	2	25463212	T	C	p.M761V	0.41	Missense	JAK2V617F	-
602	ET	-	-	-	-	-	-	-	-	MPLW515	Chr1p
603	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
604	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
605	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
606	ET	ASXL1	20	31022899	_	delC	p.W796fs*22	0.01	frameshift (+1bp)	CALR	-
607	ET	TP53	17	7579599	G	A	p.?	0.43	Splice	JAK2V617F	-
608	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
609	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
610	ET	-	-	-	-	-	-	-	-	-	-
611	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
612	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
613	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
614	ET	-	-	-	-	-	-	-	-	-	-
615	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
616	ET	MPL	1	43818306	T	G	p.Y591D	0.44	Missense	JAK2V617F	-
617	ET	DNMT3A	2	25458593	C	T	p.W860*	0.2	Nonsense	JAK2V617F	-
618	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
619	ET	-	-	-	-	-	-	-	-	CALR	-
620	ET	TET2	4	106182959	T	A	p.M1333K	0.11	Missense	CALR	Chr20
620	ET	NFE2	12	54686478	_	delG	p.R268fs*34	0.31	frameshift (+1bp)	CALR	Chr20
621	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
622	ET	-	-	-	-	-	-	-	-	JAK2V617F	-

623	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
624	ET	DNMT3A	2	25457242	C	T	p.R882H	0.32	Missense	JAK2V617F	-
625	ET	-	-	-	-	-	-	-	-	-	-
626	ET	TET2	4	106197373	C	G	p.Y1902*	0.25	Nonsense	JAK2V617F	-
626	ET	SRSF2	17	74732959	G	C	p.P95R	0.31	Missense	JAK2V617F	-
627	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
628	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
629	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
630	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
631	ET	-	-	-	-	-	-	-	-	-	-
632	ET	IDH1	2	209113112	C	T	p.R132H	0.21	Missense	JAK2V617F	Chr9
633	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
634	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
635	ET	-	-	-	-	-	-	-	-	CALR	-
636	ET	IDH1	2	209113112	C	T	p.R132H	0.4	Missense	JAK2V617F	-
636	ET	ASXL1	20	31021543	_	delTG	p.V515fs*13	0.45	frameshift (+2bp)	JAK2V617F	-
637	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
638	ET	-	-	-	-	-	-	-	-	CALR	-
639	ET	-	-	-	-	-	-	-	-	CALR	-
640	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
641	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
642	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
643	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
644	ET	-	-	-	-	-	-	-	-	-	-
645	ET	PHF6	X	133511696	_	delTGTGGCT	p.C17fs*14	0.41	frameshift (+1bp)	JAK2V617F	-
646	ET	-	-	-	-	-	-	-	-	CALR	-

647	ET	NFE2	12	54686317	_	delG	p.M322fs*5	0.17	frameshift (+1bp)	JAK2V617F	-
648	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
649	ET	DNMT3A	2	25463593	C	A	p.E697*	0.46	Nonsense	JAK2V617F	-
650	ET	ASXL1	20	31022403	_	delCACCA<13>GCGGC	p.E635fs*15	0.29	frameshift (+2bp)	JAK2V617F	-
651	ET	DNMT3A	2	25462020	C	T	p.G796D	0.35	Missense	JAK2V617F,CALR	-
652	ET	-	-	-	-	-	-	-	-	CALR	-
653	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
654	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
655	ET	-	-	-	-	-	-	-	-	CALR	-
656	ET	-	-	-	-	-	-	-	-	CALR	-
657	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
658	ET	TET2	4	106164741	C	G	p.S1203R	0.48	Missense	CALR	-
659	ET	-	-	-	-	-	-	-	-	MPLS204P	Chr18
660	ET	-	-	-	-	-	-	-	-	-	-
661	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
662	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
663	ET	-	-	-	-	-	-	-	-	CALR	-
664	ET	DNMT3A	2	25457161	A	C	p.F909C	0.16	Missense	JAK2V617F	-
664	ET	TET2	4	106156747	C	T	p.R550*	0.16	Nonsense	JAK2V617F	-
665	ET	-	-	-	-	-	-	-	-	CALR	-
666	ET	-	-	-	-	-	-	-	-	-	-
667	ET	-	-	-	-	-	-	-	-	-	-
668	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
669	ET	DNMT3A	2	25457243	G	A	p.R882C	0.37	Missense	MPLW515	-
670	ET	-	-	-	-	-	-	-	-	-	-
671	ET	NF1	17	29654554	G	A	p.R1769Q	0.46	Missense	CALR	-

672	ET	-	-	-	-	-	-	-	-	CALR	-
673	ET	PPM1D	17	58740559	_	delTTCT	p.S489fs*1	0.02	frameshift (+1bp)	JAK2V617F	-
673	ET	ASXL1	20	31022403	_	delCACCA<13>GCGGC	p.E635fs*15	0.27	frameshift (+2bp)	JAK2V617F	-
674	ET	TET2	4	106196575	_	delC	p.L1637fs*58	0.03	frameshift (+1bp)	JAK2V617F	Chr9
675	ET	-	-	-	-	-	-	-	-	-	-
676	ET	-	-	-	-	-	-	-	-	CALR	-
677	ET	-	-	-	-	-	-	-	-	-	-
678	ET	DNMT3A	2	25457259	G	T	p.D876E	0.05	Missense	JAK2V617F	Chr17
678	ET	SH2B3	12	111885301	C	T	p.R397W	0.52	Missense	JAK2V617F	Chr17
679	ET	-	-	-	-	-	-	-	-	-	-
680	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
681	ET	-	-	-	-	-	-	-	-	CALR	-
682	ET	TET2	4	106155580	_	insA	p.D162fs*9	0.21	frameshift (+2bp)	CALR	-
683	ET	TET2	4	106164071	C	G	p.P1194R	0.21	Missense	JAK2V617F	-
684	ET	SRSF2	17	74732959	G	T	p.P95H	0.33	Missense	JAK2V617F	-
684	ET	ASXL1	20	31021250	C	T	p.R417*	0.38	Nonsense	JAK2V617F	-
685	ET	TET2	4	106156304	C	G	p.S402*	0.27	Nonsense	JAK2V617F	-
686	ET	-	-	-	-	-	-	-	-	CALR	-
687	ET	-	-	-	-	-	-	-	-	CALR	-
688	ET	-	-	-	-	-	-	-	-	JAK2V617F	Chr9
689	ET	-	-	-	-	-	-	-	-	CALR	-
690	ET	-	-	-	-	-	-	-	-	-	-
691	ET	NFE2	12	54686495	_	delCTCT	p.E261fs*3	0.14	frameshift (+1bp)	JAK2V617F	-
692	ET	NFE2	12	54686575	_	delA	p.P236fs*12	0.07	frameshift (+1bp)	CALR	-
692	ET	PPM1D	17	58740540	_	delT	p.L482fs*3	0.14	frameshift (+1bp)	CALR	-
693	ET	GNB1	1	1747229	T	C	p.K57E	0.29	Missense	CALR	-

694	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
695	ET	TET2	4	106157023	C	T	p.Q642*	0.08	Nonsense	JAK2V617F	Chr9
696	ET	TET2	4	106155195	_	delA	p.K33fs*16	0.03	frameshift (+1bp)	JAK2V617F	-
697	ET	DNMT3A	2	25467448	C	T	p.G543D	0.29	Missense	JAK2V617F	-
697	ET	DNMT3A	2	25467433	A	G	p.M548T	0.31	Missense	JAK2V617F	-
697	ET	U2AF1	21	44514777	T	C	p.Q157R	0.26	Missense	JAK2V617F	-
698	ET	-	-	-	-	-	-	-	-	MPLS505N	-
699	ET	-	-	-	-	-	-	-	-	-	-
700	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
701	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
702	ET	-	-	-	-	-	-	-	-	CALR	-
703	ET	-	-	-	-	-	-	-	-	-	-
704	ET	-	-	-	-	-	-	-	-	-	-
705	ET	-	-	-	-	-	-	-	-	-	-
706	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
707	ET	DNMT3A	2	25469028	C	T	p.?	0.29	Splice	JAK2V617F	-
708	ET	-	-	-	-	-	-	-	-	-	-
709	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
710	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
711	ET	DNMT3A	2	25464544	C	T	p.V657M	0.08	Missense	JAK2V617F	-
712	ET	-	-	-	-	-	-	-	-	CALR	-
713	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
714	ET	TET2	4	106164020	T	A	p.I1177N	0.2	Missense	JAK2V617F	-
715	ET	-	-	-	-	-	-	-	-	MPLS204P	-
716	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
717	ET	IDH2	15	90631934	C	T	p.R140Q	0.1	Missense	CALR	-

718	ET	-	-	-	-	-	-	-	-	CALR	-
719	ET	TP53	17	7578190	T	C	p.Y220C	0.55	Missense	-	Chr17
720	ET	-	-	-	-	-	-	-	-	-	-
721	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
722	ET	-	-	-	-	-	-	-	-	CALR	-
723	ET	ASXL1	20	31023272	_	insA	p.P920fs*4	0.29	frameshift (+2bp)	CALR	-
724	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
725	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
726	ET	TET2	4	106197309	A	G	p.H1881R	0.11	Missense	JAK2V617F	-
727	ET	TET2	4	106196497	_	delT	p.S1611fs*5	0.16	frameshift (+1bp)	MPLW515	-
728	ET	DNMT3A	2	25470464	G	T	p.S337*	0.23	Nonsense	JAK2V617F	-
729	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
730	ET	JAK2	9	5077468	A	C	p.E627A	0.43	Missense	-	-
731	ET	TET2	4	106196705	C	T	p.Q1680*	0.06	Nonsense	JAK2V617F	-
731	ET	ASXL1	20	31022411	C	A	p.C632*	0.36	Nonsense	JAK2V617F	-
732	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
733	ET	-	-	-	-	-	-	-	-	-	-
734	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
735	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
736	ET	-	-	-	-	-	-	-	-	MPLW515	-
737	ET	ASXL1	20	31022650	_	delC	p.M713fs*12	0.13	frameshift (+1bp)	JAK2V617F	-
738	ET	-	-	-	-	-	-	-	-	MPLW515	-
739	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
740	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
741	ET	CUX1	7	101840530	_	delC	p.L614fs*25	0.4	frameshift (+1bp)	JAK2V617F	Chr14
741	ET	U2AF1	21	44514777	T	G	p.Q157P	0.44	Missense	JAK2V617F	Chr14

742	ET	-	-	-	-	-	-	-	-	CALR	Chr13
743	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
744	ET	SRSF2	17	74732959	G	T	p.P95H	0.38	Missense	MPLW515	-
745	ET	-	-	-	-	-	-	-	-	CALR	-
746	ET	TET2	4	106164791	_	delC	p.C1221fs*5	0.47	frameshift (+1bp)	MPLW515	-
747	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
748	ET	TET2	4	106157946	_	delTGCTG	p.A950fs*20	0.13	frameshift (+2bp)	JAK2V617F	-
749	ET	ASXL1	20	31022484	G	T	p.E657*	0.38	Nonsense	JAK2V617F	-
750	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
752	ET	MLL3	7	151970884	A	C	p.Y306*	0.05	Nonsense	JAK2V617F	-
753	ET	-	-	-	-	-	-	-	-	CALR	-
754	ET	-	-	-	-	-	-	-	-	CALR	-
755	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
756	ET	-	-	-	-	-	-	-	-	-	-
757	ET	-	-	-	-	-	-	-	-	CALR	-
758	ET	-	-	-	-	-	-	-	-	CALR	Chr20
759	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
760	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
761	ET	-	-	-	-	-	-	-	-	CALR	-
762	ET	TET2	4	106190798	G	T	p.R1359L	0.32	Missense	JAK2V617F	-
762	ET	TET2	4	106158248	C	A	p.S1050*	0.29	Nonsense	JAK2V617F	-
762	ET	ASXL1	20	31023356	_	insG	p.E948fs*7	0.37	frameshift (+2bp)	JAK2V617F	-
763	ET	TET2	4	106164940	G	A	p.?	0.37	Splice	JAK2V617F	-
763	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
764	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
765	ET	DNMT3A	2	25457242	C	T	p.R882H	0.08	Missense	JAK2V617F	-

766	ET	-	-	-	-	-	-	-	-	-	-
767	ET	SH2B3	12	111884663	G	A	p.?	0.5	Splice	CALR	-
768	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
769	ET	-	-	-	-	-	-	-	-	JAK2V617F	Chr9
770	ET	TET2	4	106162494	_	delAG	p.?	0.41	Splice	JAK2V617F	-
771	ET	-	-	-	-	-	-	-	-	CALR	-
772	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
773	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
774	ET	DNMT3A	2	25457176	G	A	p.P904L	0.46	Missense	CALR	-
774	ET	PPM1D	17	58740467	C	T	p.R458*	0.06	Nonsense	CALR	-
775	ET	TP53	17	7577156	C	T	p.?	0.09	Splice	CALR	-
776	ET	MPL	1	43818306	T	G	p.Y591D	0.07	Missense	JAK2V617F	-
776	ET	TET2	4	106193787	G	T	p.V1417F	0.06	Missense	JAK2V617F	-
776	ET	TP53	17	7578236	A	C	p.Y205D	0.04	Missense	JAK2V617F	-
777	ET	-	-	-	-	-	-	-	-	-	-
778	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
779	ET	MPL	1	43818310	G	A	p.R592Q	0.09	Missense	-	Chr1p
780	ET	SF3B1	2	198266834	T	C	p.K700E	0.28	Missense	JAK2V617F	-
781	ET	-	-	-	-	-	-	-	-	MPLW515	-
782	ET	MLL3	7	151970884	A	C	p.Y306*	0.05	Nonsense	CALR	-
783	ET	-	-	-	-	-	-	-	-	CALR	-
784	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
785	ET	-	-	-	-	-	-	-	-	CALR	-
786	ET	-	-	-	-	-	-	-	-	CALR	-
787	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
788	ET	CBL	11	119103398	C	T	p.Q146*	0.5	Nonsense	JAK2V617F	-

788	ET	ASXL1	20	31023000	C	T	p.Q829*	0.24	Nonsense	JAK2V617F	-
789	ET	-	-	-	-	-	-	-	-	CALR	-
790	ET	TP53	17	7577130	A	G	p.F270L	0.06	Missense	JAK2V617F	-
791	ET	TET2	4	106158018	C	A	p.C973*	0.13	Nonsense	JAK2V617F	-
792	ET	ASXL1	20	31022635	_	insC	p.Q708fs*10	0.17	frameshift (+2bp)	JAK2V617F	-
793	ET	-	-	-	-	-	-	-	-	CALR	-
794	ET	CUX1	7	101844882	G	A	p.A769T	0.54	Missense	JAK2V617F	-
795	ET	-	-	-	-	-	-	-	-	JAK2V617F,CALR	-
796	ET	ASXL1	20	31022815	_	delC	p.Q768fs*4	0.05	frameshift (+1bp)	JAK2V617F	-
797	ET	DNMT3A	2	25469646	C	T	p.?	0.31	Splice	JAK2V617F	-
798	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
799	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
800	ET	-	-	-	-	-	-	-	-	CALR	-
801	ET	-	-	-	-	-	-	-	-	-	-
802	ET	-	-	-	-	-	-	-	-	JAK2V617F	Chr9
803	ET	DNMT3A	2	25468121	C	T	p.?	0.29	Splice	JAK2V617F	-
804	ET	PPM1D	17	58740506	_	delC	p.P471fs*12	0.05	frameshift (+1bp)	CALR	-
805	ET	DNMT3A	2	25467408	C	T	p.?	0.05	Splice	JAK2V617F	-
806	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
807	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
808	ET	-	-	-	-	-	-	-	-	CALR	-
809	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
810	ET	DNMT3A	2	25457242	C	T	p.R882H	0.35	Missense	JAK2V617F	-
811	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
812	ET	-	-	-	-	-	-	-	-	CALR	-
813	ET	-	-	-	-	-	-	-	-	JAK2V617F	-

814	ET	TET2	4	106158133	_	delC	p.P1012fs*21	0.08	frameshift (+1bp)	JAK2V617F	-
814	ET	NF1	17	29670127	T	A	p.F2388Y	0.12	Missense	JAK2V617F	-
815	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
816	ET	-	-	-	-	-	-	-	-	-	Chr20
817	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
818	ET	MLL3	7	151970896	A	T	p.C302*	0.05	Nonsense	CALR	-
819	ET	-	-	-	-	-	-	-	-	-	-
820	ET	TET2	4	106196459	_	insT	p.Y1598fs*16	0.07	frameshift (+2bp)	JAK2V617F	-
821	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
822	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
823	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
824	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
825	ET	TP53	17	7577124	C	T	p.V272M	0.08	Missense	CALR	-
826	ET	-	-	-	-	-	-	-	-	-	-
827	ET	-	-	-	-	-	-	-	-	CALR	-
828	ET	ASXL1	20	31022403	_	delCACCA<13>GCGGC	p.E635fs*15	0.29	frameshift (+2bp)	JAK2V617F	-
829	ET	-	-	-	-	-	-	-	-	-	-
830	ET	DNMT3A	2	25470484	C	T	p.W330*	0.42	Nonsense	JAK2V617F	-
830	ET	TET2	4	106197285	T	C	p.I1873T	0.06	Missense	JAK2V617F	-
830	ET	TET2	4	106193779	_	delAGCT	p.Q1414fs*33	0.44	frameshift (+1bp)	JAK2V617F	-
830	ET	PTPN11	12	112924336	G	A	p.V428M	0.17	Missense	JAK2V617F	-
831	ET	RB1	13	48942685	C	T	p.R358*	0.06	Nonsense	JAK2V617F	Chr9
832	ET	-	-	-	-	-	-	-	-	JAK2V617F	Chr9
833	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
834	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
835	ET	-	-	-	-	-	-	-	-	JAK2V617F	-

836	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
837	ET	-	-	-	-	-	-	-	-	-	-
838	ET	DNMT3A	2	25458696	T	G	p.?	0.25	Splice	-	Chr12
838	ET	SF3B1	2	198266834	T	C	p.K700E	0.2	Missense	-	Chr12
839	ET	TET2	4	106156729	C	T	p.R544*	0.33	Nonsense	JAK2V617F	-
840	ET	DNMT3A	2	25463565	C	T	p.G706E	0.06	Missense	JAK2V617F	-
841	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
842	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
843	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
844	ET	-	-	-	-	-	-	-	-	-	-
845	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
846	ET	TET2	4	106157609	_	delAT	p.H839fs*6	0.51	frameshift (+2bp)	JAK2V617F	-
847	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
848	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
849	ET	-	-	-	-	-	-	-	-	CALR	-
850	ET	-	-	-	-	-	-	-	-	-	-
851	ET	-	-	-	-	-	-	-	-	JAK2V617F,CALR	-
852	ET	TET2	4	106193984	_	delG	p.E1483fs*88	0.1	frameshift (+1bp)	JAK2V617F	-
852	ET	EZH2	7	148511206	G	T	p.R566S	0.05	Missense	JAK2V617F	-
853	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
854	ET	-	-	-	-	-	-	-	-	JAK2V617F,CALR	-
855	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
856	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
857	ET	-	-	-	-	-	-	-	-	CALR	-
858	ET	-	-	-	-	-	-	-	-	CALR	-
859	ET	-	-	-	-	-	-	-	-	JAK2V617F	-

860	ET	MBD1	18	47800580	C	T	p.W374*	0.06	Nonsense	JAK2V617F	-
861	ET	TET2	4	106158075	T	A	p.C992*	0.09	Nonsense	CALR	-
862	ET	U2AF1	21	44514777	T	C	p.Q157R	0.22	Missense	JAK2V617F	-
863	ET	DNMT3A	2	25470535	C	T	p.W313*	0.12	Nonsense	JAK2V617F	-
864	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
865	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
866	ET	ASXL1	20	31022937	_	delC	p.P808fs*10	0.01	frameshift (+1bp)	JAK2V617F	-
867	ET	DNMT3A	2	25467433	A	G	p.M548T	0.37	Missense	JAK2V617F	-
868	ET	-	-	-	-	-	-	-	-	CALR	-
869	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
870	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
871	ET	-	-	-	-	-	-	-	-	-	-
872	ET	DNMT3A	2	25463310	C	T	p.G728D	0.3	Missense	CALR	-
873	ET	-	-	-	-	-	-	-	-	CALR	-
874	ET	-	-	-	-	-	-	-	-	CALR	-
875	ET	TET2	4	106157324	_	insAGCA	p.Q744fs*11	0.24	frameshift (+2bp)	JAK2V617F	-
876	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
877	ET	-	-	-	-	-	-	-	-	-	-
878	ET	-	-	-	-	-	-	-	-	CALR	-
879	ET	-	-	-	-	-	-	-	-	-	-
880	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
881	ET	TET2	4	106157527	C	T	p.Q810*	0.09	Nonsense	JAK2V617F	-
882	ET	TET2	4	106197014	C	T	p.Q1783*	0.08	Nonsense	JAK2V617F	-
882	ET	TET2	4	106156036	_	insA	p.T313fs*18	0.1	frameshift (+2bp)	JAK2V617F	-
883	ET	PPM1D	17	58740364	_	insG	p.E424fs*10	0.08	frameshift (+2bp)	CALR	-
884	ET	-	-	-	-	-	-	-	-	MPLW515	-

885	ET	SF3B1	2	198267491	C	G	p.E622D	0.4	Missense	CALR	-
886	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
887	ET	DNMT3A	2	25467448	C	T	p.G543D	0.45	Missense	CALR	-
888	ET	ASXL1	20	31022465	_	insTGGC	p.G652fs*7	0.25	frameshift (+2bp)	CALR	-
889	ET	-	-	-	-	-	-	-	-	-	-
890	ET	-	-	-	-	-	-	-	-	CALR	-
891	ET	TP53	17	7577138	C	G	p.R267P	0.05	Missense	JAK2V617F	Chr17
892	ET	-	-	-	-	-	-	-	-	-	-
893	ET	TET2	4	106190785	GC	CT	p.A1355L	0.18	Missense	CALR	-
894	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
895	ET	-	-	-	-	-	-	-	-	CALR	-
896	ET	-	-	-	-	-	-	-	-	CALR	-
897	ET	SF3B1	2	198266834	T	C	p.K700E	0.39	Missense	CALR	Chr19
897	ET	TET2	4	106164778	C	T	p.R1216*	0.44	Nonsense	CALR	Chr19
897	ET	RUNX1	21	36252852	A	G	p.?	0.33	Splice	CALR	Chr19
898	ET	-	-	-	-	-	-	-	-	-	-
899	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
900	ET	SF3B1	2	198267359	C	A	p.K666N	0.19	Missense	JAK2V617F	-
901	ET	DNMT3A	2	25463286	C	T	p.R736H	0.08	Missense	MPLW515	-
902	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
903	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
904	ET	TP53	17	7578525	G	T	p.C135*	0.3	Nonsense	JAK2V617F	Chr17
905	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
906	ET	-	-	-	-	-	-	-	-	CALR	-
907	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
908	ET	IDH2	15	90631934	C	T	p.R140Q	0.23	Missense	JAK2V617F	-

909	ET	-	-	-	-	-	-	-	-	-	-
910	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
911	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
912	ET	SF3B1	2	198267361	T	G	p.K666Q	0.47	Missense	JAK2V617F	-
913	ET	ZRSR2	X	15840936	G	A	p.W340*	0.09	Nonsense	JAK2V617F	-
914	ET	-	-	-	-	-	-	-	-	CALR	-
915	ET	SRSF2	17	74732959	G	T	p.P95H	0.32	Missense	MPLW515	-
916	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
917	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
918	ET	-	-	-	-	-	-	-	-	JAK2V617F	Chr9
919	ET	-	-	-	-	-	-	-	-	JAK2V617F	Chr9
920	ET	DNMT3A	2	25463298	A	G	p.F732S	0.06	Missense	CALR	-
921	ET	TET2	4	106157317	C	T	p.Q740*	0.16	Nonsense	JAK2V617F	-
922	ET	-	-	-	-	-	-	-	-	-	-
923	ET	SF3B1	2	198267361	T	G	p.K666Q	0.2	Missense	MPLW515	-
924	ET	TET2	4	106158349	C	T	p.Q1084*	0.05	Nonsense	JAK2V617F	-
925	ET	-	-	-	-	-	-	-	-	CALR	Chr20
926	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
927	ET	TET2	4	106155921	_	delC	p.N275fs*18	0.01	frameshift (+1bp)	CALR	-
928	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
929	ET	-	-	-	-	-	-	-	-	MPLW515	-
930	ET	TET2	4	106157123	C	A	p.S675*	0.05	Nonsense	-	-
931	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
932	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
933	ET	-	-	-	-	-	-	-	-	-	-
934	ET	-	-	-	-	-	-	-	-	JAK2V617F	-

935	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
936	ET	MBD1	18	47800074	G	A	p.Q436*	0.31	Nonsense	JAK2V617F	-
937	ET	-	-	-	-	-	-	-	-	-	-
938	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
939	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
940	ET	TET2	4	106164764	G	A	p.C1211Y	0.42	Missense	JAK2V617F	-
940	ET	TET2	4	106155439	C	T	p.Q114*	0.39	Nonsense	JAK2V617F	-
940	ET	PPM1D	17	58740546	T	G	p.L484*	0.38	Nonsense	JAK2V617F	-
940	ET	TP53	17	7578554	A	C	p.Y126D	0.07	Missense	JAK2V617F	-
941	ET	NF1	17	29562683	C	T	p.Q1255*	0.18	Nonsense	JAK2V617F	-
942	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
943	ET	-	-	-	-	-	-	-	-	-	-
944	ET	-	-	-	-	-	-	-	-	-	-
945	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
946	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
947	ET	-	-	-	-	-	-	-	-	CALR	-
948	ET	ASXL1	20	31022592	C	T	p.R693*	0.36	Nonsense	JAK2V617F	Chr14,20
949	ET	-	-	-	-	-	-	-	-	-	-
950	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
951	ET	-	-	-	-	-	-	-	-	CALR	-
952	ET	-	-	-	-	-	-	-	-	CALR	-
953	ET	PPM1D	17	58740529	C	A	p.C478*	0.16	Nonsense	JAK2V617F,CALR	-
954	ET	NFE2	12	54686380	_	delCCGCTCCAGCTC	p.E297_R300delELER	0.03	Inframe	CALR	Chr5
954	ET	PPM1D	17	58740475	_	delT	p.N460fs*5	0.41	frameshift (+1bp)	CALR	Chr5
955	ET	-	-	-	-	-	-	-	-	-	-
956	ET	-	-	-	-	-	-	-	-	JAK2V617F	-

957	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
958	ET	-	-	-	-	-	-	-	-	-	-
959	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
960	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
961	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
962	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
963	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
964	ET	-	-	-	-	-	-	-	-	-	-
965	ET	ASXL1	20	31021211	C	T	p.R404*	0.08	Nonsense	JAK2V617F	-
966	ET	TET2	4	106190861	A	C	p.H1380P	0.12	Missense	JAK2V617F	Chr9
967	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
968	ET	NFE2	12	54686429	C	T	p.R284H	0.17	Missense	JAK2V617F	-
968	ET	PPM1D	17	58740509	G	T	p.E472*	0.06	Nonsense	JAK2V617F	-
969	ET	-	-	-	-	-	-	-	-	JAK2V617F	Chr9
970	ET	TP53	17	7578463	C	T	p.R156H	0.48	Missense	JAK2V617F	-
971	ET	-	-	-	-	-	-	-	-	CALR	-
972	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
973	ET	-	-	-	-	-	-	-	-	CALR	-
974	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
975	ET	SRSF2	17	74732959	G	T	p.P95H	0.38	Missense	JAK2V617F	-
975	ET	ASXL1	20	31021274	_	insT	p.Y425fs*13	0.45	frameshift (+2bp)	JAK2V617F	-
976	ET	DNMT3A	2	25457243	G	A	p.R882C	0.39	Missense	JAK2V617F	-
976	ET	TET2	4	106180931	G	A	p.?	0.06	Splice	JAK2V617F	-
976	ET	TET2	4	106157615	_	insA	p.H839fs*7	0.16	frameshift (+2bp)	JAK2V617F	-
976	ET	CBL	11	119167639	_	delT	p.V683fs*19	0.2	frameshift (+1bp)	JAK2V617F	-
977	ET	-	-	-	-	-	-	-	-	CALR	-

978	ET	-	-	-	-	-	-	-	-	-	-
979	ET	-	-	-	-	-	-	-	-	-	-
980	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
981	ET	-	-	-	-	-	-	-	-	CALR	-
982	ET	DNMT3A	2	25467449	C	A	p.G543C	0.27	Missense	JAK2V617F	-
983	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
984	ET	-	-	-	-	-	-	-	-	MPLW515	-
985	ET	-	-	-	-	-	-	-	-	CALR	-
986	ET	-	-	-	-	-	-	-	-	CALR	-
987	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
988	ET	-	-	-	-	-	-	-	-	CALR	-
989	ET	-	-	-	-	-	-	-	-	-	-
990	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
991	ET	-	-	-	-	-	-	-	-	JAK2V617F	Chr17
992	ET	ASXL1	20	31023153	_	insA	p.T880fs*2	0.16	frameshift (+2bp)	CALR	-
993	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
994	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
995	ET	-	-	-	-	-	-	-	-	-	Chr12
996	ET	-	-	-	-	-	-	-	-	-	-
997	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
998	ET	-	-	-	-	-	-	-	-	CALR	-
999	ET	-	-	-	-	-	-	-	-	MPLW515	-
1000	ET	-	-	-	-	-	-	-	-	CALR	-
1001	ET	-	-	-	-	-	-	-	-	CALR	-
1002	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
1003	ET	ASXL1	20	31023646	_	insC	p.K1045fs*5	0.18	frameshift (+2bp)	JAK2V617F	-

1004	ET	-	-	-	-	-	-	-	-	CALR	-
1005	ET	TET2	4	106156414	_	delA	p.N439fs*8	0.33	frameshift (+1bp)	CALR	-
1006	ET	-	-	-	-	-	-	-	-	-	-
1007	ET	DNMT3A	2	25463248	G	A	p.R749C	0.37	Missense	CALR	-
1008	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
1009	ET	-	-	-	-	-	-	-	-	CALR	-
1011	ET	-	-	-	-	-	-	-	-	-	-
1012	ET	TET2	4	106157404	C	T	p.Q769*	0.4	Nonsense	CALR	-
1012	ET	TET2	4	106190782	A	G	p.R1354G	0.09	Missense	CALR	-
1013	ET	TET2	4	106180816	_	delG	p.G1282fs*81	0.05	frameshift (+1bp)	JAK2V617F	-
1013	ET	TET2	4	106190860	C	T	p.H1380Y	0.08	Missense	JAK2V617F	-
1014	ET	EZH2	7	148526858	A	C	p.L149R	0.09	Missense	CALR	-
1014	ET	EZH2	7	148507470	C	T	p.V662M	0.27	Missense	CALR	-
1014	ET	TP53	17	7576539	A	G	p.*347Q	0.07	Stop loss	CALR	-
1014	ET	TP53	17	7577556	C	A	p.C242F	0.09	Missense	CALR	-
1014	ET	ASXL1	20	31022403	_	delCACCA<13>GCGGC	p.E635fs*15	0.22	frameshift (+2bp)	CALR	-
1015	ET	CUX1	7	101877462	G	A	p.W1188*	0.05	Nonsense	JAK2V617F	-
1015	ET	U2AF1	21	44514777	T	C	p.Q157R	0.38	Missense	JAK2V617F	-
1016	ET	-	-	-	-	-	-	-	-	-	-
1017	ET	-	-	-	-	-	-	-	-	-	-
1018	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
1019	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
1020	ET	-	-	-	-	-	-	-	-	-	-
1021	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
1022	ET	-	-	-	-	-	-	-	-	-	-
1023	ET	-	-	-	-	-	-	-	-	-	-

1024	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
1025	ET	-	-	-	-	-	-	-	-	-	-
1026	ET	CBL	11	119148919	T	C	p.L380P	0.07	Missense	JAK2V617F	Chr18
1027	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
1028	ET	NF1	17	29533395	A	T	p.?	0.51	Splice	JAK2V617F	-
1029	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
1030	ET	-	-	-	-	-	-	-	-	JAK2V617F	Chr17
1031	ET	-	-	-	-	-	-	-	-	-	-
1032	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
1033	ET	-	-	-	-	-	-	-	-	JAK2V617F	Chr9
1034	ET	-	-	-	-	-	-	-	-	JAK2V617F	Chr18
1035	ET	DNMT3A	2	25464538	G	C	p.R659G	0.05	Missense	CALR	-
1035	ET	TET2	4	106190843	G	A	p.C1374Y	0.15	Missense	CALR	-
1036	ET	-	-	-	-	-	-	-	-	-	-
1037	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
1038	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
1039	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
1040	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
1042	ET	-	-	-	-	-	-	-	-	CALR	-
1043	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
1044	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
1045	ET	-	-	-	-	-	-	-	-	CALR	-
1046	ET	-	-	-	-	-	-	-	-	-	-
1047	ET	-	-	-	-	-	-	-	-	MPLW515	-
1048	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
1049	ET	-	-	-	-	-	-	-	-	JAK2V617F	-

1050	ET	-	-	-	-	-	-	-	-	MPLW515	-
1051	ET	-	-	-	-	-	-	-	-	-	-
1052	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
1053	ET	PPM1D	17	58740546	_	insA	p.L484fs*5	0.14	frameshift (+2bp)	JAK2V617F	-
1054	ET	PPM1D	17	58740526	_	delT	p.C478fs*5	0.19	frameshift (+1bp)	JAK2V617F	-
1055	ET	DNMT3A	2	25463212	T	C	p.M761V	0.09	Missense	JAK2V617F	Chr5
1056	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
1057	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
1059	ET	-	-	-	-	-	-	-	-	-	-
1060	ET	DNMT3A	2	25469041	_	delC	p.E473fs*178	0.06	frameshift (+1bp)	JAK2V617F	-
1061	ET	-	-	-	-	-	-	-	-	-	-
1062	ET	-	-	-	-	-	-	-	-	-	-
1063	ET	-	-	-	-	-	-	-	-	-	-
1064	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
1065	ET	-	-	-	-	-	-	-	-	CALR	-
1066	ET	-	-	-	-	-	-	-	-	-	-
1067	ET	-	-	-	-	-	-	-	-	-	-
1068	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
1069	ET	-	-	-	-	-	-	-	-	CALR	-
1070	ET	-	-	-	-	-	-	-	-	-	-
1071	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
1072	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
1073	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
1074	ET	-	-	-	-	-	-	-	-	JAK2V617F	Chr1p
1075	ET	TET2	4	106196213	C	T	p.R1516*	0.09	Nonsense	CALR	-
1076	ET	-	-	-	-	-	-	-	-	JAK2V617F	-

1077	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
1078	ET	-	-	-	-	-	-	-	-	-	-
1079	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
1080	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
1081	ET	-	-	-	-	-	-	-	-	-	-
1082	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
1083	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
1084	ET	-	-	-	-	-	-	-	-	-	-
1085	ET	-	-	-	-	-	-	-	-	-	-
1086	ET	DNMT3A	2	25459832	_	delC	p.E817fs*8	0.27	frameshift (+1bp)	-	-
1087	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
1088	ET	TET2	4	106156729	C	T	p.R544*	0.07	Nonsense	JAK2V617F	-
1089	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
1090	ET	-	-	-	-	-	-	-	-	MPLW515	-
1091	ET	-	-	-	-	-	-	-	-	CALR	-
1092	ET	-	-	-	-	-	-	-	-	CALR	-
1093	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
1094	ET	MPL	1	43805052	G	A	p.E168K	0.54	Missense	-	-
1095	ET	-	-	-	-	-	-	-	-	CALR	-
1096	ET	-	-	-	-	-	-	-	-	CALR	-
1097	ET	-	-	-	-	-	-	-	-	-	-
1098	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
1099	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
1100	ET	-	-	-	-	-	-	-	-	CALR	-
1101	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
1102	ET	-	-	-	-	-	-	-	-	JAK2V617F	-

1103	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
1104	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
1105	ET	JAK2	9	5078360	A	G	p.R683G	0.42	Missense	-	-
1106	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
1107	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
1108	ET	-	-	-	-	-	-	-	-	CALR	-
1109	ET	-	-	-	-	-	-	-	-	-	-
1110	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
1111	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
1112	ET	-	-	-	-	-	-	-	-	CALR	-
1113	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
1114	ET	U2AF1	21	44514777	T	C	p.Q157R	0.16	Missense	CALR	-
1115	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
1116	ET	-	-	-	-	-	-	-	-	-	-
1117	ET	-	-	-	-	-	-	-	-	-	-
1118	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
1119	ET	-	-	-	-	-	-	-	-	CALR	-
1120	ET	-	-	-	-	-	-	-	-	-	-
1121	ET	-	-	-	-	-	-	-	-	-	-
1122	ET	-	-	-	-	-	-	-	-	-	-
1123	ET	-	-	-	-	-	-	-	-	-	-
1124	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
1125	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
1126	ET	-	-	-	-	-	-	-	-	CALR	-
1127	ET	-	-	-	-	-	-	-	-	JAK2V617F	Chr9
1128	ET	-	-	-	-	-	-	-	-	JAK2V617F	-

1129	ET	-	-	-	-	-	-	-	-	CALR	-
1130	ET	MLL3	7	151882660	_	delC	p.E1689fs*28	0.09	frameshift (+1bp)	CALR	-
1131	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
1132	ET	-	-	-	-	-	-	-	-	-	-
1133	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
1134	ET	-	-	-	-	-	-	-	-	CALR	-
1135	ET	-	-	-	-	-	-	-	-	JAK2V617F	Chr5
1136	ET	-	-	-	-	-	-	-	-	-	-
1137	ET	-	-	-	-	-	-	-	-	CALR	-
1138	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
1139	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
1140	ET	TET2	4	106156069	C	T	p.Q324*	0.31	Nonsense	JAK2V617F	-
1140	ET	ASXL1	20	31022403	_	delCACCA<13>GCGGC	p.E635fs*15	0.03	frameshift (+2bp)	JAK2V617F	-
1140	ET	ASXL1	20	31022592	C	T	p.R693*	0.34	Nonsense	JAK2V617F	-
1141	ET	-	-	-	-	-	-	-	-	CALR	-
1142	ET	-	-	-	-	-	-	-	-	CALR	-
1143	ET	-	-	-	-	-	-	-	-	JAK2V617F	Chr9
1144	ET	-	-	-	-	-	-	-	-	CALR	-
1145	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
1146	ET	-	-	-	-	-	-	-	-	CALR	-
1147	ET	TET2	4	106155521	T	C	p.V141A	0.33	Missense	CALR	-
1148	ET	DNMT3A	2	25457243	G	A	p.R882C	0.1	Missense	JAK2V617F	-
1149	ET	CBL	11	119148925	A	G	p.K382R	0.18	Missense	JAK2V617F	-
1149	ET	ASXL1	20	31022229	T	C	p.?	0.6	Splice	JAK2V617F	-
1150	ET	-	-	-	-	-	-	-	-	CALR	-
1151	ET	-	-	-	-	-	-	-	-	CALR	-

1152	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
1153	ET	-	-	-	-	-	-	-	-	JAK2V617F	Chr9
1154	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
1155	ET	TET2	4	106164860	A	G	p.K1243R	0.48	Missense	JAK2V617F	-
1156	ET	-	-	-	-	-	-	-	-	-	-
1157	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
1158	ET	TET2	4	106194061	_	delC	p.A1508fs*63	0.16	frameshift (+1bp)	CALR	-
1158	ET	TET2	4	106164791	_	delC	p.C1221fs*5	0.06	frameshift (+1bp)	CALR	-
1159	ET	-	-	-	-	-	-	-	-	CALR	-
1160	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
1161	ET	-	-	-	-	-	-	-	-	CALR	-
1162	ET	-	-	-	-	-	-	-	-	CALR	-
1163	ET	-	-	-	-	-	-	-	-	-	-
1164	ET	MLL3	7	151882660	_	delC	p.E1689fs*28	0.07	frameshift (+1bp)	-	-
1164	ET	SH2B3	12	111856059	_	insC	p.A38fs*85	0.24	frameshift (+2bp)	-	-
1165	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
1166	ET	-	-	-	-	-	-	-	-	CALR	-
1167	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
1168	ET	-	-	-	-	-	-	-	-	CALR	-
1169	ET	-	-	-	-	-	-	-	-	MPLW515	-
1170	ET	-	-	-	-	-	-	-	-	-	-
1171	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
1172	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
1173	ET	MLL3	7	151970884	A	C	p.Y306*	0.04	Nonsense	-	-
1174	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
1175	ET	-	-	-	-	-	-	-	-	-	-

1176	ET	-	-	-	-	-	-	-	-	-	-
1177	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
1178	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
1179	ET	-	-	-	-	-	-	-	-	CALR	-
1180	ET	-	-	-	-	-	-	-	-	CALR	-
1181	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
1182	ET	-	-	-	-	-	-	-	-	CALR	-
1183	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
1184	ET	-	-	-	-	-	-	-	-	CALR	-
1185	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
1186	ET	TET2	4	106196329	_	delAG	p.E1555fs*22	0.04	frameshift (+2bp)	JAK2V617F	-
1187	ET	-	-	-	-	-	-	-	-	-	-
1188	ET	-	-	-	-	-	-	-	-	CALR	-
1189	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
1190	ET	CBL	11	119148978	A	T	p.M400L	0.53	Missense	CALR	-
1190	ET	PPM1D	17	58740544	_	insT	p.L484fs*5	0.05	frameshift (+2bp)	CALR	-
1190	ET	BCOR	X	39911579	A	C	p.F1684C	0.05	Missense	CALR	-
1191	ET	-	-	-	-	-	-	-	-	-	-
1192	ET	-	-	-	-	-	-	-	-	CALR	-
1193	ET	-	-	-	-	-	-	-	-	CALR	-
1194	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
1195	ET	-	-	-	-	-	-	-	-	-	-
1196	ET	-	-	-	-	-	-	-	-	CALR	-
1197	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
1198	ET	-	-	-	-	-	-	-	-	CALR	-
1199	ET	-	-	-	-	-	-	-	-	JAK2V617F	-

1200	ET	-	-	-	-	-	-	-	-	-	-
1201	ET	-	-	-	-	-	-	-	-	-	-
1202	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
1203	ET	-	-	-	-	-	-	-	-	CALR	-
1204	ET	-	-	-	-	-	-	-	-	CALR	-
1205	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
1206	ET	-	-	-	-	-	-	-	-	CALR	-
1207	ET	-	-	-	-	-	-	-	-	MPLW515	-
1208	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
1209	ET	IDH2	15	90631934	C	T	p.R140Q	0.63	Missense	JAK2V617F	-
1209	ET	SRSF2	17	74732960	_	insGGC	p.R94_P95insR	0.3	Inframe	JAK2V617F	-
1209	ET	ASXL1	20	31023918	C	T	p.Q1135*	0.45	Nonsense	JAK2V617F	-
1210	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
1211	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
1212	ET	-	-	-	-	-	-	-	-	-	-
1213	ET	-	-	-	-	-	-	-	-	-	-
1214	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
1215	ET	SF3B1	2	198266482	_	delTTCATTTCC	p.E783_K785delEMK	0.27	Inframe	JAK2V617F	-
1215	ET	PTPN11	12	112926270	C	T	p.T468M	0.41	Missense	JAK2V617F	-
1216	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
1217	ET	DNMT3A	2	25457242	C	T	p.R882H	0.32	Missense	JAK2V617F	-
1218	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
1219	ET	-	-	-	-	-	-	-	-	-	-
1220	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
1221	ET	-	-	-	-	-	-	-	-	-	-
1222	ET	-	-	-	-	-	-	-	-	JAK2V617F	-

1223	ET	-	-	-	-	-	-	-	-	MPLW515	-
1224	ET	DNMT3A	2	25458595	A	G	p.W860R	0.07	Missense	-	-
1224	ET	JAK2	9	5089701	C	T	p.R867W	0.47	Missense	-	-
1225	ET	-	-	-	-	-	-	-	-	CALR	-
1226	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
1227	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
1228	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
1229	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
1230	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
1231	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
1232	ET	TET2	4	106197552	C	T	p.P1962L	0.46	Missense	-	-
1232	ET	ASXL1	20	31021276	C	G	p.Y425*	0.05	Nonsense	-	-
1233	ET	-	-	-	-	-	-	-	-	CALR	-
1234	ET	-	-	-	-	-	-	-	-	-	-
1235	ET	TET2	4	106182914	_	insGTAT	p.?	0.32	Splice	CALR	-
1235	ET	MLL3	7	151970896	A	T	p.C302*	0.04	Nonsense	CALR	-
1236	ET	-	-	-	-	-	-	-	-	CALR	-
1237	ET	-	-	-	-	-	-	-	-	-	-
1238	ET	JAK2	9	5089702	G	A	p.R867Q	0.49	Missense	-	-
1239	ET	-	-	-	-	-	-	-	-	CALR	-
1240	ET	-	-	-	-	-	-	-	-	-	-
1241	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
1242	ET	-	-	-	-	-	-	-	-	MPLS204P	-
1243	ET	-	-	-	-	-	-	-	-	CALR	-
1244	ET	DNMT3A	2	25463297	_	delA	p.F732fs*47	0.35	frameshift (+1bp)	CALR	-
1245	ET	-	-	-	-	-	-	-	-	CALR	-

1246	ET	-	-	-	-	-	-	-	-	CALR	-
1247	ET	-	-	-	-	-	-	-	-	-	-
1248	ET	TET2	4	106180929	T	G	p.?	0.5	Splice	JAK2V617F	-
1249	ET	-	-	-	-	-	-	-	-	CALR	-
1250	ET	TET2	4	106196537	C	T	p.Q1624*	0.08	Nonsense	CALR	-
1251	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
1252	ET	-	-	-	-	-	-	-	-	-	-
1253	ET	-	-	-	-	-	-	-	-	-	-
1254	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
1255	ET	-	-	-	-	-	-	-	-	CALR	-
1256	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
1257	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
1258	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
1259	ET	-	-	-	-	-	-	-	-	CALR	Chr17
1260	ET	-	-	-	-	-	-	-	-	CALR	-
1261	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
1262	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
1263	ET	-	-	-	-	-	-	-	-	CALR	-
1264	ET	KIT	4	55595599	C	T	p.H697Y	0.44	Missense	JAK2V617F	-
1265	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
1266	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
1267	ET	-	-	-	-	-	-	-	-	CALR	-
1268	ET	KIT	4	55602733	G	A	p.V852I	0.46	Missense	JAK2V617F	-
1269	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
1270	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
1271	ET	ASXL1	20	31023153	_	insA	p.T880fs*2	0.3	frameshift (+2bp)	CALR	-

1272	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
1273	ET	-	-	-	-	-	-	-	-	-	-
1274	ET	-	-	-	-	-	-	-	-	MPLS505N	-
1275	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
1276	ET	-	-	-	-	-	-	-	-	CALR	-
1277	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
1278	ET	-	-	-	-	-	-	-	-	CALR	-
1279	ET	-	-	-	-	-	-	-	-	-	-
1280	ET	-	-	-	-	-	-	-	-	MPLW515	-
1281	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
1282	ET	MPL	1	43812147	G	C	p.E338Q	0.48	Missense	-	-
1283	ET	TET2	4	106157769	_	insA	p.Q891fs*10	0.16	frameshift (+2bp)	JAK2V617F	-
1284	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
1285	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
1286	ET	MLL3	7	151970896	A	T	p.C302*	0.04	Nonsense	CALR	-
1287	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
1288	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
1289	ET	-	-	-	-	-	-	-	-	CALR	-
1290	ET	-	-	-	-	-	-	-	-	-	-
1291	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
1292	ET	CBL	11	119148978	A	T	p.M400L	0.51	Missense	CALR	-
1293	ET	IDH2	15	90631934	C	T	p.R140Q	0.47	Missense	JAK2V617F	Chr9
1294	ET	EZH2	7	148508788	C	T	p.V626M	0.38	Missense	JAK2V617F	-
1294	ET	TP53	17	7577563	T	C	p.S240G	0.23	Missense	JAK2V617F	-
1295	ET	EZH2	7	148506443	C	T	p.R690H	0.41	Missense	JAK2V617F	-
1295	ET	PPM1D	17	58740460	_	delG	p.I456fs*1	0.08	frameshift (+1bp)	JAK2V617F	-

1295	ET	PPM1D	17	58740642	C	G	p.S516*	0.14	Nonsense	JAK2V617F	-
1296	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
1297	ET	-	-	-	-	-	-	-	-	CALR	-
1298	ET	DNMT3A	2	25469541	C	T	p.W409*	0.17	Nonsense	JAK2V617F	-
1299	ET	DNMT3A	2	25463302	A	C	p.F731V	0.44	Missense	MPLW515	-
1299	ET	ASXL1	20	31023437	C	G	p.Y974*	0.39	Nonsense	MPLW515	-
1300	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
1301	ET	-	-	-	-	-	-	-	-	-	-
1302	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
1303	ET	-	-	-	-	-	-	-	-	CALR	-
1304	ET	-	-	-	-	-	-	-	-	-	-
1305	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
1306	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
1307	ET	-	-	-	-	-	-	-	-	-	-
1308	ET	-	-	-	-	-	-	-	-	CALR	-
1309	ET	NFE2	12	54687044	_	delG	p.P79fs*32	0.17	frameshift (+1bp)	JAK2V617F	-
1310	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
1311	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
1312	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
1313	ET	-	-	-	-	-	-	-	-	-	-
1314	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
1315	ET	-	-	-	-	-	-	-	-	CALR	-
1316	ET	MLL3	7	151882660	_	delC	p.E1689fs*28	0.05	frameshift (+1bp)	CALR	-
1317	ET	-	-	-	-	-	-	-	-	CALR	-
1318	ET	-	-	-	-	-	-	-	-	CALR	-
1319	ET	-	-	-	-	-	-	-	-	JAK2V617F	-

1320	ET	NFE2	12	54686495	_	delCTCT	p.E261fs*3	0.24	frameshift (+1bp)	JAK2V617F	-
1321	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
1322	ET	-	-	-	-	-	-	-	-	CALR	Chr20
1323	ET	-	-	-	-	-	-	-	-	JAK2V617F	-
1324	ET	-	-	-	-	-	-	-	-	-	-
1325	ET	-	-	-	-	-	-	-	-	JAK2V617F	Chr9
1326	ET	-	-	-	-	-	-	-	-	JAK2V617F	Chr9
1327	ET	-	-	-	-	-	-	-	-	CALR	-
1328	MF	TET2	4	106157244	_	delCT	p.S716fs*6	0.22	frameshift (+2bp)	JAK2V617F	-
1328	MF	EZH2	7	148507433	A	G	p.L674S	0.27	Missense	JAK2V617F	-
1328	MF	KRAS	12	25378562	C	G	p.A146P	0.16	Missense	JAK2V617F	-
1328	MF	ZRSR2	X	15827338	_	delA	p.T152fs*13	0.52	frameshift (+1bp)	JAK2V617F	-
1329	MF	TET2	4	106156153	_	delT	p.F352fs*20	0.41	frameshift (+1bp)	JAK2V617F	Chr9
1329	MF	TET2	4	106197353	A	G	p.R1896G	0.48	Missense	JAK2V617F	Chr9
1329	MF	CBL	11	119148922	G	A	p.C381Y	0.3	Missense	JAK2V617F	Chr9
1329	MF	SRSF2	17	74732959	G	T	p.P95H	0.14	Missense	JAK2V617F	Chr9
1329	MF	ASXL1	20	31022614	_	insGAG	p.Y700delins*	0.2	Nonsense	JAK2V617F	Chr9
1330	MF	TET2	4	106197285	T	A	p.I1873N	0.97	Missense	MPLW515	Chr1p,4
1330	MF	SRSF2	17	74732959	G	T	p.P95H	0.48	Missense	MPLW515	Chr1p,4
1331	MF	TP53	17	7578212	G	A	p.R213*	0.61	Nonsense	CALR	Chr17
1331	MF	ASXL1	20	31022782	_	delACCAG<15>CCCCT	p.Q757fs*7	0.33	frameshift (+1bp)	CALR	Chr17
1331	MF	ASXL1	20	31022592	C	T	p.R693*	0.1	Nonsense	CALR	Chr17
1332	MF	TET2	4	106190861	A	C	p.H1380P	0.49	Missense	JAK2V617F	Chr9,14
1333	MF	MPL	1	43818405	C	G	p.H624D	0.32	Missense	JAK2V617F	-
1333	MF	TET2	4	106158419	C	A	p.S1107*	0.33	Nonsense	JAK2V617F	-
1333	MF	CUX1	7	101459338	C	T	p.Q10*	0.38	Nonsense	JAK2V617F	-

1333	MF	PTPN11	12	112888172	A	G	p.Y63C	0.24	Missense	JAK2V617F	-
1333	MF	ASXL1	20	31022983	T	A	p.L823*	0.26	Nonsense	JAK2V617F	-
1333	MF	U2AF1	21	44514770	_	insCTCATA	p.E159_M160insYE	0.21	Inframe	JAK2V617F	-
1334	MF	ASXL1	20	31023591	G	T	p.G1026*	0.05	Nonsense	JAK2V617F,CALR	Chr20
1335	MF	-	-	-	-	-	-	-	-	JAK2V617F	Chr9
1336	MF	NFE2	12	54686495	_	delCTCT	p.E261fs*3	0.16	frameshift (+1bp)	JAK2V617F	Chr9,20
1337	MF	-	-	-	-	-	-	-	-	JAK2V617F	-
1338	MF	SETBP1	18	42531919	G	C	p.G872R	0.08	Missense	MPLW515	Chr1p
1338	MF	ASXL1	20	31022403	_	delCACCA<13>GCGGC	p.E635fs*15	0.08	frameshift (+2bp)	MPLW515	Chr1p
1338	MF	ASXL1	20	31023717	C	T	p.R1068*	0.17	Nonsense	MPLW515	Chr1p
1338	MF	U2AF1	21	44514777	T	C	p.Q157R	0.48	Missense	MPLW515	Chr1p
1339	MF	GNAS	20	57484421	G	A	p. R201H	0.46	Missense	JAK2V617F	Chr9
1339	MF	ZRSR2	X	15809095	_	delG	p.R27fs*11	0.71	frameshift (+1bp)	JAK2V617F	Chr9
1340	MF	TET2	4	106155778	G	T	p.E227*	0.44	Nonsense	JAK2V617F	Chr9
1340	MF	MLL3	7	151970896	A	T	p.C302*	0.06	Nonsense	JAK2V617F	Chr9
1341	MF	DNMT3A	2	25457242	C	T	p.R882H	0.08	Missense	JAK2V617F	Chr9
1342	MF	NRAS	1	115258747	C	T	p.G12D	0.1	Missense	JAK2V617F	Chr9
1342	MF	EZH2	7	148506444	G	T	p.R690S	0.47	Missense	JAK2V617F	Chr9
1342	MF	PPM1D	17	58740691	_	delA	p.N533fs*6	0.48	frameshift (+1bp)	JAK2V617F	Chr9
1342	MF	ASXL1	20	31022550	G	T	p.G679*	0.41	Nonsense	JAK2V617F	Chr9
1342	MF	ASXL1	20	31022817	C	T	p.Q768*	0.15	Nonsense	JAK2V617F	Chr9
1342	MF	ZRSR2	X	15827389	C	T	p.R169*	0.97	Nonsense	JAK2V617F	Chr9
1343	MF	ASXL1	20	31022403	_	delCACCA<13>GCGGC	p.E635fs*15	0.12	frameshift (+2bp)	JAK2V617F	Chr9,13,14
1344	MF	-	-	-	-	-	-	-	-	CALR	-
1345	MF	TET2	4	106197357	T	G	p.I1897S	0.09	Missense	JAK2V617F	-
1345	MF	ASXL1	20	31022288	C	G	p.Y591*	0.45	Nonsense	JAK2V617F	-

1345	MF	U2AF1	21	44514777	T	G	p.Q157P	0.27	Missense	JAK2V617F	-
1346	MF	-	-	-	-	-	-	-	-	CALR	Chr19
1347	MF	-	-	-	-	-	-	-	-	-	-
1348	MF	-	-	-	-	-	-	-	-	CALR	-
1349	MF	-	-	-	-	-	-	-	-	JAK2V617F	Chr9
1350	MF	-	-	-	-	-	-	-	-	JAK2V617F	Chr9
1351	MF	-	-	-	-	-	-	-	-	MPLW515	Chr1p
1352	MF	TP53	17	7578535	T	C	p.K132R	0.38	Missense	JAK2V617F	Chr5,9,17,20
1352	MF	GNAS	20	57484597	G	T	p.Q227H	0.45	Missense	JAK2V617F	Chr5,9,17,20
1353	MF	-	-	-	-	-	-	-	-	CALR	Chr7
1354	MF	U2AF1	21	44514777	T	G	p.Q157P	0.25	Missense	JAK2V617F	-
1355	MF	-	-	-	-	-	-	-	-	CALR	-
1356	MF	SF3B1	2	198267359	C	A	p.K666N	0.46	Missense	JAK2V617F	-
1356	MF	TET2	4	106156306	C	T	p.Q403*	0.47	Nonsense	JAK2V617F	-
1357	MF	TET2	4	106196684	_	delC	p.P1673fs*22	0.94	frameshift (+1bp)	JAK2V617F	Chr4,7,9
1358	MF	-	-	-	-	-	-	-	-	-	-
1359	MF	ASXL1	20	31022258	_	delACCCTGGGTGGTTA	p.P582fs*32	0.13	frameshift (+2bp)	JAK2V617F	-
1360	MF	-	-	-	-	-	-	-	-	CALR	-
1361	MF	TET2	4	106155526	G	A	p.D143N	0.52	Missense	JAK2V617F	-
1361	MF	TET2	4	106158427	A	T	p.K1110*	0.61	Nonsense	JAK2V617F	-
1361	MF	TET2	4	106196217	_	insT	p.S1518fs*60	0.33	frameshift (+2bp)	JAK2V617F	-
1362	MF	-	-	-	-	-	-	-	-	JAK2V617F	Chr9
1363	MF	SF3B1	2	198266834	T	C	p.K700E	0.41	Missense	CALR,MPLS204P	-
1363	MF	CBL	11	119148931	G	A	p.C384Y	0.3	Missense	CALR,MPLS204P	-
1364	MF	SF3B1	2	198267359	C	A	p.K666N	0.41	Missense	CALR	-
1365	MF	-	-	-	-	-	-	-	-	JAK2V617F	-

1366	MF	DNMT3A	2	25463533	_	delTGACGATGGAG	p.S714fs*16	0.3	frameshift (+2bp)	CALR	-
1367	MF	-	-	-	-	-	-	-	-	MPLS204P	Chr1p
1368	MF	NRAS	1	115258747	C	T	p.G12D	0.39	Missense	CALR	-
1368	MF	ASXL1	20	31022847	C	T	p.Q778*	0.48	Nonsense	CALR	-
1369	MF	-	-	-	-	-	-	-	-	JAK2V617F	-
1370	MF	TET2	4	106164913	C	T	p.R1261C	0.34	Missense	CALR	-
1370	MF	ASXL1	20	31022937	_	delC	p.P808fs*10	0.4	frameshift (+1bp)	CALR	-
1371	MF	-	-	-	-	-	-	-	-	JAK2V617F	-
1372	MF	-	-	-	-	-	-	-	-	CALR	-
1373	MF	EZH2	7	148506172	A	G	p.F729S	0.42	Missense	CALR	Chr7
1374	MF	ASXL1	20	31022697	G	T	p.E728*	0.47	Nonsense	MPLW515	-
1375	MF	ASXL1	20	31022403	_	delCACCA<13>GCGGC	p.E635fs*15	0.2	frameshift (+2bp)	CALR	-
1375	MF	U2AF1	21	44514777	T	C	p.Q157R	0.74	Missense	CALR	-
1376	MF	-	-	-	-	-	-	-	-	CALR	-
1377	MF	-	-	-	-	-	-	-	-	JAK2V617F	Chr9
1378	MF	-	-	-	-	-	-	-	-	JAK2V617F	Chr9
1379	MF	TET2	4	106155471	_	insA	p.N124fs*12	0.38	frameshift (+2bp)	CALR	-
1379	MF	NFE2	12	54686430	G	A	p.R284C	0.5	Missense	CALR	-
1380	MF	EZH2	7	148507475	C	T	p.G660E	0.24	Missense	MPLW515	Chr1p
1380	MF	ASXL1	20	31024148	_	delCTCC	p.L1213fs*3	0.49	frameshift (+1bp)	MPLW515	Chr1p
1381	MF	DNMT3A	2	25457242	C	T	p.R882H	0.57	Missense	JAK2V617F	-
1381	MF	SF3B1	2	198266834	T	C	p.K700E	0.49	Missense	JAK2V617F	-
1381	MF	TET2	4	106155749	_	delC	p.V218fs*32	0.47	frameshift (+1bp)	JAK2V617F	-
1382	MF	NF1	17	29559200	T	G	p.F1103V	0.53	Missense	CALR	Chr1p
1382	MF	ASXL1	20	31022403	_	delCACCA<13>GCGGC	p.E635fs*15	0.16	frameshift (+2bp)	CALR	Chr1p
1383	MF	ASXL1	20	31022592	C	T	p.R693*	0.29	Nonsense	JAK2V617F	Chr9

1384	MF	-	-	-	-	-	-	-	-	CALR	Chr13
1385	MF	-	-	-	-	-	-	-	-	CALR	-
1386	MF	RUNX1	21	36252849	C	A	p.?	0.29	Splice	CALR	Chr7
1387	MF	STAG2	X	123164963	G	C	p.K92N	0.13	Missense	CALR	-
1388	MF	-	-	-	-	-	-	-	-	CALR	-
1389	MF	-	-	-	-	-	-	-	-	CALR	-
1390	MF	ASXL1	20	31023269	_	delCATAC	p.P920fs*2	0.36	frameshift (+2bp)	JAK2V617F	-
1391	MF	PPM1D	17	58740537	_	delC	p.L482fs*1	0.51	frameshift (+1bp)	JAK2V617F	Chr9
1391	MF	U2AF1	21	44514777	T	G	p.Q157P	0.35	Missense	JAK2V617F	Chr9
1392	MF	-	-	-	-	-	-	-	-	JAK2V617F,MPLW515	-
1393	MF	TET2	4	106193885	_	delT	p.R1451fs*7	0.1	frameshift (+1bp)	CALR	-
1394	MF	-	-	-	-	-	-	-	-	JAK2V617F	Chr9,18
1395	MF	EZH2	7	148508788	C	T	p.V626M	0.52	Missense	JAK2V617F	-
1396	MF	TET2	4	106193892	C	T	p.R1452*	0.97	Nonsense	JAK2V617F	Chr4
1396	MF	EZH2	7	148543629	C	T	p.W60*	0.43	Nonsense	JAK2V617F	Chr4
1396	MF	KRAS	12	25398285	C	T	p.G12S	0.45	Missense	JAK2V617F	Chr4
1396	MF	ASXL1	20	31021290	C	G	p.S430*	0.47	Nonsense	JAK2V617F	Chr4
1397	MF	-	-	-	-	-	-	-	-	JAK2V617F	-
1398	MF	TET2	4	106156876	C	T	p.Q593*	0.61	Nonsense	JAK2V617F	Chr4,9
1398	MF	EZH2	7	148506477	C	T	p.V679M	0.52	Missense	JAK2V617F	Chr4,9
1399	MF	TET2	4	106196975	_	delC	p.P1770fs*50	0.05	frameshift (+1bp)	JAK2V617F	Chr20
1399	MF	ASXL1	20	31024148	_	delCTCC	p.L1213fs*3	0.05	frameshift (+1bp)	JAK2V617F	Chr20
1400	MF	NFE2	12	54687140	G	T	p.S47*	0.56	Nonsense	JAK2V617F	-
1401	MF	U2AF1	21	44514777	T	G	p.Q157P	0.3	Missense	JAK2V617F	-
1401	MF	ZRSR2	X	15818076	_	delG	p.R68fs*10	0.14	frameshift (+1bp)	JAK2V617F	-
1402	MF	SRSF2	17	74732959	G	T	p.P95H	0.56	Missense	JAK2V617F	Chr9

1403	MF	TET2	4	106190796	_	delC	p.R1359fs*4	0.05	frameshift (+1bp)	CALR	-
1403	MF	EZH2	7	148523590	C	T	p.R288Q	0.47	Missense	CALR	-
1403	MF	ASXL1	20	31022592	C	T	p.R693*	0.54	Nonsense	CALR	-
1404	MF	EZH2	7	148506477	C	T	p.V679M	0.96	Missense	-	Chr7
1404	MF	ASXL1	20	31022238	C	T	p.Q575*	0.5	Nonsense	-	Chr7
1405	MF	-	-	-	-	-	-	-	-	JAK2V617F	-
1406	MF	-	-	-	-	-	-	-	-	CALR	-
1407	MF	RUNX1	21	36252877	C	T	p.R162K	0.13	Missense	-	-
1408	MF	-	-	-	-	-	-	-	-	CALR	-
1409	MF	TET2	4	106155858	_	delCAGT	p.S254fs*38	0.9	frameshift (+1bp)	JAK2V617F	Chr9
1410	MF	TP53	17	7577548	C	T	p.G245S	0.61	Missense	JAK2V617F	Chr9,17
1410	MF	ASXL1	20	31022615	T	A	p.Y700*	0.47	Nonsense	JAK2V617F	Chr9,17
1410	MF	U2AF1	21	44514777	T	C	p.Q157R	0.33	Missense	JAK2V617F	Chr9,17
1411	MF	TP53	17	7577555	G	C	p.C242W	0.58	Missense	JAK2V617F	-
1412	MF	CBL	11	119148880	A	C	p.Q367P	0.36	Missense	-	-
1412	MF	SRSF2	17	74732959	G	C	p.P95R	0.43	Missense	-	-
1413	MF	EZH2	7	148544387	C	A	p.G2C	0.24	Missense	JAK2V617F	-
1414	MF	-	-	-	-	-	-	-	-	-	-
1415	MF	TET2	4	106190797	C	T	p.R1359C	0.19	Missense	-	Chr7,11
1415	MF	CBL	11	119149251	G	A	p.R420Q	0.91	Missense	-	Chr7,11
1415	MF	ASXL1	20	31021643	_	insT	p.R549fs*2	0.31	frameshift (+2bp)	-	Chr7,11
1415	MF	BCOR	X	39923680	_	delTT	p.K1137fs*4	0.15	frameshift (+2bp)	-	Chr7,11
1416	MF	-	-	-	-	-	-	-	-	JAK2V617F	-
1417	MF	-	-	-	-	-	-	-	-	JAK2V617F	-
1418	MF	-	-	-	-	-	-	-	-	JAK2V617F	-
1419	MF	-	-	-	-	-	-	-	-	JAK2V617F	-

1420	MF	SF3B1	2	198267359	C	G	p.K666N	0.4	Missense	JAK2V617F	-
1421	MF	RB1	13	48923124	_	delTAAAAGT	p.K192fs*7	0.11	frameshift (+1bp)	MPLW515	Chr1p,18
1422	MF	-	-	-	-	-	-	-	-	JAK2V617F	Chr9
1423	MF	-	-	-	-	-	-	-	-	JAK2V617F	-
1424	MF	-	-	-	-	-	-	-	-	CALR	-
1425	MF	SF3B1	2	198266720	G	A	p.H738Y	0.14	Missense	-	-
1426	MF	KRAS	12	25398285	C	G	p.G12R	0.23	Missense	CALR	-
1427	MF	NRAS	1	115258747	C	T	p.G12D	0.43	Missense	JAK2V617F	-
1427	MF	TET2	4	106196213	C	T	p.R1516*	0.45	Nonsense	JAK2V617F	-
1427	MF	TET2	4	106197360	C	T	p.S1898F	0.41	Missense	JAK2V617F	-
1427	MF	ASXL1	20	31022287	_	insA	p.Y591fs*1	0.48	frameshift (+2bp)	JAK2V617F	-
1428	MF	NRAS	1	115258748	C	G	p.G12R	0.19	Missense	JAK2V617F	-
1428	MF	IDH1	2	209113112	C	T	p.R132H	0.43	Missense	JAK2V617F	-
1428	MF	ASXL1	20	31022592	C	T	p.R693*	0.36	Nonsense	JAK2V617F	-
1429	MF	SF3B1	2	198268392	C	A	p.D546Y	0.15	Missense	-	-
1430	MF	-	-	-	-	-	-	-	-	JAK2V617F	-
1431	MF	CBL	11	119148873	_	insTdelAGGAA<4>ATGA A	p.?	0.24	Splice	JAK2V617F	-
1431	MF	ASXL1	20	31021211	C	T	p.R404*	0.32	Nonsense	JAK2V617F	-
1432	MF	DNMT3A	2	25457243	G	A	p.R882C	0.48	Missense	-	-
1432	MF	EZH2	7	148506443	C	T	p.R690H	0.27	Missense	-	-
1432	MF	U2AF1	21	44524456	G	A	p.S34F	0.25	Missense	-	-
1432	MF	STAG2	X	123220440	C	T	p.R1033*	0.14	Nonsense	-	-
1433	MF	SF3B1	2	198257839	C	T	p.E1205K	0.15	Missense	MPLW515	Chr1p
1433	MF	SF3B1	2	198267359	C	G	p.K666N	0.12	Missense	MPLW515	Chr1p
1434	MF	ASXL1	20	31023408	C	T	p.R965*	0.37	Nonsense	JAK2V617F	-
1434	MF	U2AF1	21	44514777	T	G	p.Q157P	0.45	Missense	JAK2V617F	-

1435	MF	TET2	4	106164778	C	T	p.R1216*	0.37	Nonsense	JAK2V617F	-
1435	MF	SRSF2	17	74732959	G	T	p.P95H	0.5	Missense	JAK2V617F	-
1436	MF	DNMT3A	2	25463248	G	T	p.R749S	0.29	Missense	JAK2V617F	Chr9
1436	MF	TET2	4	106158311	C	G	p.A1071G	0.26	Missense	JAK2V617F	Chr9
1437	MF	SRSF2	17	74732959	G	T	p.P95H	0.48	Missense	JAK2V617F	Chr9
1438	MF	SRSF2	17	74732960	G	C	p.P95A	0.5	Missense	JAK2V617F	Chr17,20
1439	MF	-	-	-	-	-	-	-	-	CALR	-
1440	MF	EZH2	7	148526901	C	T	p.G135R	0.17	Missense	JAK2V617F	-
1440	MF	ASXL1	20	31022403	_	delCACCA<13>GCGGC	p.E635fs*15	0.46	frameshift (+2bp)	JAK2V617F	-
1441	MF	-	-	-	-	-	-	-	-	JAK2V617F	-
1442	MF	DNMT3A	2	25457243	G	A	p.R882C	0.47	Missense	JAK2V617F	Chr9
1442	MF	TET2	4	106180782	_	insT	p.C1271fs*29	0.12	frameshift (+2bp)	JAK2V617F	Chr9
1443	MF	RUNX1	21	36252877	C	T	p.R162K	0.23	Missense	JAK2V617F	-
1443	MF	RUNX1	21	36231859	_	insCAAGG	p.T176fs*2	0.06	frameshift (+1bp)	JAK2V617F	-
1444	MF	-	-	-	-	-	-	-	-	JAK2V617F	-
1445	MF	TET2	4	106196346	_	insATTCTGCTTC	p.G1564fs*17	0.2	frameshift (+2bp)	JAK2V617F	-
1445	MF	TET2	4	106158123	_	insG	p.Q1009fs*9	0.43	frameshift (+2bp)	JAK2V617F	-
1446	MF	DNMT3A	2	25457243	G	A	p.R882C	0.33	Missense	CALR	-
1447	MF	EZH2	7	148507425	C	G	p.D677H	1	Missense	JAK2V617F	Chr7
1447	MF	ASXL1	20	31022287	_	insA	p.Y591fs*1	0.68	frameshift (+2bp)	JAK2V617F	Chr7
1448	MF	ZRSR2	X	15827389	C	T	p.R169*	0.75	Nonsense	CALR	-
1448	MF	ZRSR2	X	15809095	_	delG	p.R27fs*11	0.08	frameshift (+1bp)	CALR	-
1449	MF	-	-	-	-	-	-	-	-	JAK2V617F	-
1450	MF	EZH2	7	148516705	G	A	p.Q328*	0.42	Nonsense	JAK2V617F	-
1450	MF	ASXL1	20	31022288	C	A	p.Y591*	0.31	Nonsense	JAK2V617F	-
1451	MF	SF3B1	2	198266834	T	C	p.K700E	0.54	Missense	CALR	-

1451	MF	TET2	4	106156531	_	delG	p.E478fs*8	0.06	frameshift (+1bp)	CALR	-
1451	MF	TP53	17	7579599	G	A	p.?	0.58	Splice	CALR	-
1451	MF	TP53	17	7577547	C	T	p.G245D	0.39	Missense	CALR	-
1451	MF	ASXL1	20	31022319	G	A	p.E602K	0.19	Missense	CALR	-
1452	MF	-	-	-	-	-	-	-	-	JAK2V617F	Chr9
1453	MF	TET2	4	106180865	G	A	p.C1298Y	0.51	Missense	JAK2V617F	-
1454	MF	-	-	-	-	-	-	-	-	CALR	-
1455	MF	SF3B1	2	198267359	C	G	p.K666N	0.5	Missense	JAK2V617F	-
1456	MF	CBL	11	119149217	C	T	p.?	0.26	Splice	-	-
1456	MF	TP53	17	7576593	G	T	p.?	0.16	Splice	-	-
1457	MF	-	-	-	-	-	-	-	-	JAK2V617F	Chr9
1458	MF	MPL	1	43818310	G	A	p.R592Q	0.52	Missense	JAK2V617F	Chr17
1458	MF	TET2	4	106180931	G	A	p.?	0.49	Splice	JAK2V617F	Chr17
1458	MF	CBL	11	119142446	C	T	p.R149*	0.34	Nonsense	JAK2V617F	Chr17
1458	MF	SRSF2	17	74732959	G	C	p.P95R	0.78	Missense	JAK2V617F	Chr17
1458	MF	ASXL1	20	31023436	_	insA	p.Y974fs*1	0.47	frameshift (+2bp)	JAK2V617F	Chr17
1459	MF	SRSF2	17	74732959	G	A	p.P95L	0.22	Missense	JAK2V617F	-
1460	MF	ASXL1	20	31023398	_	delTG	p.V962fs*7	0.42	frameshift (+2bp)	JAK2V617F	-
1461	MF	-	-	-	-	-	-	-	-	-	-
1462	MF	-	-	-	-	-	-	-	-	MPLW515	Chr1p
1463	MF	ASXL1	20	31022288	_	insA	p.Y591fs*1	0.5	frameshift (+2bp)	JAK2V617F	-
1464	MF	ASXL1	20	31022287	_	insA	p.Y591fs*1	0.49	frameshift (+2bp)	JAK2V617F	-
1465	MF	SF3B1	2	198266834	T	C	p.K700E	0.4	Missense	JAK2V617F	-
1465	MF	TET2	4	106193778	C	G	p.Q1414E	0.63	Missense	JAK2V617F	-
1466	MF	-	-	-	-	-	-	-	-	JAK2V617F	-
1467	MF	-	-	-	-	-	-	-	-	JAK2V617F	-

1468	MF	TET2	4	106197325	_	delAAAGA	p.K1887fs*3	0.08	frameshift (+2bp)	CALR	-
1468	MF	EZH2	7	148508781	C	T	p.G628D	0.34	Missense	CALR	-
1468	MF	ASXL1	20	31022902	G	A	p.W796*	0.56	Nonsense	CALR	-
1468	MF	GNAS	20	57484421	G	A	p.R201H	0.43	Missense	CALR	-
1468	MF	ZRSR2	X	15833814	_	delAT	p.H191fs*12	0.73	frameshift (+2bp)	CALR	-
1469	MF	-	-	-	-	-	-	-	-	JAK2V617F	Chr9
1470	MF	SRSF2	17	74732959	G	T	p.P95H	0.8	Missense	JAK2V617F	-
1471	MF	NRAS	1	115258747	C	T	p.G12D	0.6	Missense	JAK2V617F	-
1471	MF	SRSF2	17	74732959	G	T	p.P95H	0.38	Missense	JAK2V617F	-
1472	MF	CBL	11	119148973	A	G	p.H398R	0.23	Missense	MPLW515	Chr1p
1472	MF	IDH2	15	90631934	C	T	p.R140Q	0.57	Missense	MPLW515	Chr1p
1472	MF	ASXL1	20	31022988	A	T	p.K825*	0.49	Nonsense	MPLW515	Chr1p
1473	MF	-	-	-	-	-	-	-	-	JAK2V617F	-
1474	MF	-	-	-	-	-	-	-	-	JAK2V617F	-
1475	MF	NRAS	1	115258747	C	T	p.G12D	0.39	Missense	-	-
1475	MF	TET2	4	106157968	_	delTT	p.L957fs*14	0.5	frameshift (+2bp)	-	-
1475	MF	GNAS	20	57484420	C	T	p.R201C	0.44	Missense	-	-
1476	MF	-	-	-	-	-	-	-	-	CALR	-
1477	MF	-	-	-	-	-	-	-	-	JAK2V617F	Chr9
1478	MF	-	-	-	-	-	-	-	-	JAK2V617F	-
1479	MF	TET2	4	106197387	T	C	p.M1907T	0.11	Missense	JAK2V617F	Chr9
1480	MF	NF1	17	29661945	C	T	p.R1968*	0.14	Nonsense	-	-
1481	MF	U2AF1	21	44514777	T	G	p.Q157P	0.57	Missense	JAK2V617F	-
1482	MF	-	-	-	-	-	-	-	-	JAK2V617F	Chr9
1483	MF	PTPN11	12	112888301	A	C	p.D106A	0.42	Missense	JAK2V617F	Chr9
1484	MF	-	-	-	-	-	-	-	-	JAK2V617F	-

1485	MF	-	-	-	-	-	-	-	-	-	-
1486	MF	NRAS	1	115258747	C	T	p.G12D	0.4	Missense	-	-
1486	MF	GATA2	3	128200137	T	C	p.K390E	0.58	Missense	-	-
1486	MF	GNAS	20	57484421	G	A	p.R201H	0.41	Missense	-	-
1487	MF	ASXL1	20	31022624	_	delT	p.N703fs*22	0.29	frameshift (+1bp)	CALR	-
1488	MF	RB1	13	48942685	C	T	p.R358*	0.39	Nonsense	-	-
1488	MF	RB1	13	49039374	C	T	p.R787*	0.45	Nonsense	-	-
1488	MF	ASXL1	20	31022319	G	T	p.E602*	0.53	Nonsense	-	-
1489	MF	PTPN11	12	112915532	A	G	p.M311V	0.39	Missense	JAK2V617F	Chr9,14
1489	MF	ASXL1	20	31022403	_	delCACCA<13>GCGGC	p.E635fs*15	0.4	frameshift (+2bp)	JAK2V617F	Chr9,14
1490	MF	TP53	17	7579353	C	T	p.G112S	0.38	Missense	JAK2V617F	Chr9,18
1491	MF	-	-	-	-	-	-	-	-	JAK2V617F	Chr9
1492	MF	-	-	-	-	-	-	-	-	JAK2V617F	-
1493	MF	SF3B1	2	198266834	T	C	p.K700E	0.17	Missense	JAK2V617F	-
1494	MF	-	-	-	-	-	-	-	-	JAK2V617F	-
1495	MF	-	-	-	-	-	-	-	-	JAK2V617F	Chr9
1496	MF	KRAS	12	25398285	C	T	p.G12S	0.38	Missense	JAK2V617F	-
1496	MF	ASXL1	20	31022403	_	delCACCA<13>GCGGC	p.E635fs*15	0.28	frameshift (+2bp)	JAK2V617F	-
1497	MF	-	-	-	-	-	-	-	-	JAK2V617F	Chr9
1498	MF	SRSF2	17	74732959	G	A	p.P95L	0.77	Missense	JAK2V617F	-
1499	MF	-	-	-	-	-	-	-	-	JAK2V617F	Chr20
1500	MF	IDH1	2	209103945	G	A	p.A335V	0.22	Missense	JAK2V617F	-
1500	MF	U2AF1	21	44524456	G	A	p.S34F	0.14	Missense	JAK2V617F	-
1501	MF	-	-	-	-	-	-	-	-	JAK2V617F	-
1502	MF	-	-	-	-	-	-	-	-	JAK2V617F	Chr9
1503	MF	NRAS	1	115258747	C	T	p.G12D	0.42	Missense	-	-

1503	MF	TET2	4	106164914	G	C	p.R1261P	0.47	Missense	-	-
1503	MF	TET2	4	106157302	_	delAGTTCACA	p.S736fs*15	0.32	frameshift (+2bp)	-	-
1503	MF	EZH2	7	148526820	C	T	p.E162K	0.37	Missense	-	-
1504	MF	ZRSR2	X	15840988	_	delG	p.K359fs*>130	0.25	frameshift (+1bp)	JAK2V617F	-
1505	MF	-	-	-	-	-	-	-	-	CALR	-
1506	MF	-	-	-	-	-	-	-	-	JAK2V617F	-
1507	MF	TET2	4	106193885	_	delTT	p.F1450fs*27	0.57	frameshift (+2bp)	JAK2V617F	-
1507	MF	TET2	4	106156069	C	T	p.Q324*	0.53	Nonsense	JAK2V617F	-
1507	MF	CBL	11	119149248	G	A	p.C419Y	0.42	Missense	JAK2V617F	-
1507	MF	CBL	11	119148958	T	C	p.I393T	0.16	Missense	JAK2V617F	-
1507	MF	CBL	11	119155763	C	T	p.R506*	0.21	Nonsense	JAK2V617F	-
1507	MF	ASXL1	20	31023077	_	delTGAT	p.D855fs*11	0.38	frameshift (+1bp)	JAK2V617F	-
1508	MF	U2AF1	21	44514777	T	C	p.Q157R	0.51	Missense	JAK2V617F	-
1509	MF	-	-	-	-	-	-	-	-	JAK2V617F	-
1510	MF	-	-	-	-	-	-	-	-	-	-
1511	MF	RUNX1	21	36231774	G	A	p.R204*	0.41	Nonsense	CALR	Chr7
1512	MF	-	-	-	-	-	-	-	-	JAK2V617F	Chr9
1513	MF	NRAS	1	115258747	C	T	p.G12D	0.38	Missense	-	-
1513	MF	STAG2	X	123224609	_	delT	p.Y1155fs*5	0.96	frameshift (+1bp)	-	-
1514	MF	ASXL1	20	31022712	C	T	p.Q733*	0.13	Nonsense	JAK2V617F	-
1514	MF	U2AF1	21	44514777	T	G	p.Q157P	0.15	Missense	JAK2V617F	-
1515	MF	IDH1	2	209113112	C	T	p.R132H	0.33	Missense	JAK2V617F,CALR	Chr20
1516	MF	TET2	4	106164862	_	delCT	p.Y1245fs*22	0.04	frameshift (+2bp)	JAK2V617F	Chr9
1517	MF	-	-	-	-	-	-	-	-	CALR	-
1518	MF	EZH2	7	148523590	C	T	p.R288Q	0.29	Missense	JAK2V617F	Chr8,9
1518	MF	ASXL1	20	31021250	C	T	p.R417*	0.51	Nonsense	JAK2V617F	Chr8,9

1519	MF	ASXL1	20	31023379	_	delG	p.S955fs*29	0.35	frameshift (+1bp)	JAK2V617F	-
1519	MF	U2AF1	21	44514777	T	C	p.Q157R	0.49	Missense	JAK2V617F	-
1519	MF	ZRSR2	X	15833942	C	T	p.Q234*	0.86	Nonsense	JAK2V617F	-
1520	MF	DNMT3A	2	25457242	C	T	p.R882H	0.06	Missense	JAK2V617F	-
1520	MF	DNMT3A	2	25463184	G	A	p.S770L	0.14	Missense	JAK2V617F	-
1521	MF	GATA2	3	128200131	_	delATGAAGAAGGAAins TTC	p.M388_E391delinF	0.23	Inframe	JAK2V617F	Chr1q,9
1522	MF	NF1	17	29556918	_	insCTA	p.L972_D973ins*	0.05	Nonsense	JAK2V617F	-
1523	MF	-	-	-	-	-	-	-	-	CALR	-
1524	MF	ASXL1	20	31022662	_	delC	p.A716fs*9	0.06	frameshift (+1bp)	CALR	-
1524	MF	U2AF1	21	44514777	T	G	p.Q157P	0.4	Missense	CALR	-
1525	MF	SRSF2	17	74732959	G	T	p.P95H	0.38	Missense	JAK2V617F	Chr9
1526	MF	TET2	4	106197285	T	C	p.I1873T	0.49	Missense	JAK2V617F	-
1526	MF	SRSF2	17	74732959	G	T	p.P95H	0.5	Missense	JAK2V617F	-
1526	MF	ASXL1	20	31023261	A	T	p.R916*	0.44	Nonsense	JAK2V617F	-
1527	MF	TET2	4	106182906	C	G	p.?	0.47	Splice	JAK2V617F	-
1527	MF	CBL	11	119148990	T	C	p.C404R	0.28	Missense	JAK2V617F	-
1527	MF	CBL	11	119103370	_	delA	p.E138fs*14	0.29	frameshift (+1bp)	JAK2V617F	-
1527	MF	ASXL1	20	31023408	C	T	p.R965*	0.38	Nonsense	JAK2V617F	-
1527	MF	U2AF1	21	44514777	T	C	p.Q157R	0.51	Missense	JAK2V617F	-
1527	MF	ZRSR2	X	15838385	C	T	p.R295*	0.99	Nonsense	JAK2V617F	-
1528	MF	ASXL1	20	31022277	C	T	p.Q588*	0.37	Nonsense	JAK2V617F	-
1528	MF	U2AF1	21	44514777	T	G	p.Q157P	0.29	Missense	JAK2V617F	-
1529	MF	IDH1	2	209113112	C	T	p.R132H	0.51	Missense	JAK2V617F	Chr9
1529	MF	TP53	17	7578406	C	T	p.R175H	0.22	Missense	JAK2V617F	Chr9
1529	MF	U2AF1	21	44514777	T	G	p.Q157P	0.46	Missense	JAK2V617F	Chr9
1530	MF	CBL	11	119148919	T	C	p.L380P	0.88	Missense	JAK2V617F	Chr7,11

1530	MF	SRSF2	17	74732959	G	T	p.P95H	0.55	Missense	JAK2V617F	Chr7,11
1530	MF	ASXL1	20	31022403	_	delCACCA<13>GCGGC	p.E635fs*15	0.39	frameshift (+2bp)	JAK2V617F	Chr7,11
1531	MF	MBD1	18	47803329	G	T	p.P89T	0.2	Missense	JAK2V617F	-
1532	MF	ASXL1	20	31022937	_	delC	p.P808fs*10	0.31	frameshift (+1bp)	JAK2V617F	-
1532	MF	U2AF1	21	44514777	T	G	p.Q157P	0.43	Missense	JAK2V617F	-
1533	MF	-	-	-	-	-	-	-	-	JAK2V617F	Chr9
1534	MF	DNMT3A	2	25457242	C	T	p.R882H	0.5	Missense	JAK2V617F	Chr9
1535	MF	IDH2	15	90631934	C	T	p.R140Q	0.34	Missense	JAK2V617F	Chr9
1535	MF	SRSF2	17	74732959	G	A	p.P95L	0.33	Missense	JAK2V617F	Chr9
1536	MF	ZRSR2	X	15833838	_	delC	p.L200fs*38	0.27	frameshift (+1bp)	CALR	-
1537	MF	TET2	4	106158491	_	insTTTCC	p.P1131fs*8	0.33	frameshift (+1bp)	JAK2V617F	Chr9
1537	MF	TET2	4	106196964	_	delACAGT<7>CGGGC	p.S1767fs*16	0.44	frameshift (+2bp)	JAK2V617F	Chr9
1538	MF	SF3B1	2	198266834	T	C	p.K700E	0.41	Missense	CALR	Chr20
1539	MF	ASXL1	20	31024339	C	G	p.S1275*	0.41	Nonsense	CALR	-
1540	MF	TET2	4	106196347	_	delT	p.S1561fs*10	0.42	frameshift (+1bp)	CALR	-
1541	MF	-	-	-	-	-	-	-	-	JAK2V617F	Chr9
1542	MF	-	-	-	-	-	-	-	-	JAK2V617F	-
1543	MF	-	-	-	-	-	-	-	-	-	-
1544	MF	TP53	17	7577114	C	T	p.C275Y	0.22	Missense	JAK2V617F	-
1545	MF	TET2	4	106158463	_	insCT	p.P1123fs*15	0.17	frameshift (+1bp)	JAK2V617F	Chr9
1546	MF	DNMT3A	2	25467518	_	delA	p.C520fs*131	0.3	frameshift (+1bp)	JAK2V617F	Chr9
1546	MF	DNMT3A	2	25463184	G	A	p.S770L	0.31	Missense	JAK2V617F	Chr9
1546	MF	TET2	4	106197285	T	C	p.I1873T	0.13	Missense	JAK2V617F	Chr9
1546	MF	EZH2	7	148507485	C	G	p.D657H	0.34	Missense	JAK2V617F	Chr9
1547	MF	TET2	4	106180824	_	delCTT	p.F1285delF	0.3	Inframe	MPLW515	-
1548	MF	DNMT3A	2	25457242	C	T	p.R882H	0.46	Missense	JAK2V617F	Chr9

1549	MF	DNMT3A	2	25457272	_	insTATAGTGG	p.D876fs*8	0.28	frameshift (+1bp)	JAK2V617F	-
1549	MF	ASXL1	20	31022600	_	insACTAC	p.L697fs*8	0.06	frameshift (+1bp)	JAK2V617F	-
1549	MF	ASXL1	20	31023269	_	delCATAC	p.P920fs*2	0.08	frameshift (+2bp)	JAK2V617F	-
1549	MF	ASXL1	20	31023153	_	insTA	p.T880fs*1	0.05	frameshift (+1bp)	JAK2V617F	-
1549	MF	U2AF1	21	44514777	T	C	p.Q157R	0.47	Missense	JAK2V617F	-
1550	MF	EZH2	7	148515067	_	insT	p.D382fs*5	0.29	frameshift (+2bp)	JAK2V617F	Chr9
1550	MF	ASXL1	20	31022402	_	delTC	p.H630fs*4	0.47	frameshift (+2bp)	JAK2V617F	Chr9
1551	MF	-	-	-	-	-	-	-	-	CALR	-
1552	MF	-	-	-	-	-	-	-	-	CALR	-
1553	MF	SRSF2	17	74732959	G	C	p.P95R	0.41	Missense	JAK2V617F	Chr9
1554	MF	U2AF1	21	44524456	G	T	p.S34Y	0.34	Missense	JAK2V617F	Chr14
1555	MF	-	-	-	-	-	-	-	-	JAK2V617F	-
1556	MF	PHF6	X	133511698	T	A	p.C17*	0.96	Nonsense	JAK2V617F	Chr9
1557	MF	GATA2	3	128204753	G	A	p.R230C	0.23	Missense	MPLW515	Chr1p
1558	MF	-	-	-	-	-	-	-	-	CALR	-
1559	MF	TET2	4	106182958	_	insA	p.M1333fs*6	0.59	frameshift (+2bp)	JAK2V617F	Chr9
1559	MF	ZRSR2	X	15841123	_	delA	p.R403fs*>86	0.75	frameshift (+1bp)	JAK2V617F	Chr9
1560	MF	-	-	-	-	-	-	-	-	JAK2V617F	Chr9,14
1561	MF	TET2	4	106156569	_	insA	p.Q491fs*13	0.45	frameshift (+2bp)	JAK2V617F	Chr9
1561	MF	SRSF2	17	74732959	G	T	p.P95H	0.52	Missense	JAK2V617F	Chr9
1562	MF	TET2	4	106156035	_	insA	p.T313fs*18	0.32	frameshift (+2bp)	JAK2V617F	-
1563	MF	EZH2	7	148506443	C	T	p.R690H	0.14	Missense	CALR	-
1563	MF	RB1	13	48955389	_	delC	p.T502fs*17	0.42	frameshift (+1bp)	CALR	-
1563	MF	ASXL1	20	31022712	C	T	p.Q733*	0.47	Nonsense	CALR	-
1564	MF	-	-	-	-	-	-	-	-	JAK2V617F	-
1565	MF	-	-	-	-	-	-	-	-	JAK2V617F	Chr18

1566	MF	TET2	4	106180784	G	C	p.C1271S	0.15	Missense	JAK2V617F	-
1566	MF	TET2	4	106197003	_	delC	p.L1780fs*40	0.35	frameshift (+1bp)	JAK2V617F	-
1566	MF	TET2	4	106156594	_	delC	p.P499fs*34	0.05	frameshift (+1bp)	JAK2V617F	-
1566	MF	CBL	11	119142478	_	delC	p.H160fs*6	0.17	frameshift (+1bp)	JAK2V617F	-
1567	MF	-	-	-	-	-	-	-	-	JAK2V617F	-
1568	MF	IDH1	2	209104604	G	A	p.T325M	0.14	Missense	JAK2V617F	Chr9
1568	MF	TET2	4	106180784	G	A	p.C1271Y	0.1	Missense	JAK2V617F	Chr9
1568	MF	TET2	4	106164913	C	T	p.R1261C	0.63	Missense	JAK2V617F	Chr9
1568	MF	SRSF2	17	74732959	G	A	p.P95L	0.29	Missense	JAK2V617F	Chr9
1568	MF	ASXL1	20	31021439	G	T	p.E480*	0.39	Nonsense	JAK2V617F	Chr9
1569	MF	SRSF2	17	74732959	G	C	p.P95R	0.35	Missense	-	-
1569	MF	ASXL1	20	31022839	T	G	p.L775*	0.5	Nonsense	-	-
1570	MF	TP53	17	7578389	G	A	p.R181C	0.17	Missense	JAK2V617F	Chr5
1571	MF	TET2	4	106164061	C	T	p.Q1191*	0.42	Nonsense	JAK2V617F	-
1571	MF	CUX1	7	101848438	_	delTG	p.V1041fs*17	0.31	frameshift (+2bp)	JAK2V617F	-
1571	MF	SRSF2	17	74732959	G	C	p.P95R	0.55	Missense	JAK2V617F	-
1571	MF	ASXL1	20	31023479	_	delCT	p.S989fs*1	0.35	frameshift (+2bp)	JAK2V617F	-
1572	MF	-	-	-	-	-	-	-	-	-	-
1573	MF	DNMT3A	2	25470925	_	delTC	p.D279fs*1	1	frameshift (+2bp)	CALR	Chr7,12,19
1573	MF	TET2	4	106197020	A	T	p.K1785*	0.5	Nonsense	CALR	Chr7,12,19
1574	MF	NFE2	12	54686430	G	A	p.R284C	0.13	Missense	JAK2V617F	Chr9
1575	MF	ASXL1	20	31023608	_	delG	p.D1032fs*15	0.29	frameshift (+1bp)	JAK2V617F	-
1576	MF	JAK2	9	5081734	_	delTAT	p.L816delL	0.36	Inframe	JAK2V617F	-
1577	MF	-	-	-	-	-	-	-	-	CALR	-
1578	MF	-	-	-	-	-	-	-	-	-	-
1579	MF	-	-	-	-	-	-	-	-	JAK2V617F	-

1580	MF	SF3B1	2	198267359	C	A	p.K666N	0.39	Missense	JAK2V617F	-
1581	MF	-	-	-	-	-	-	-	-	JAK2V617F	Chr9
1582	MF	-	-	-	-	-	-	-	-	-	-
1583	MF	TET2	4	106197061	_	delA	p.K1799fs*21	0.5	frameshift (+1bp)	JAK2V617F	Chr9
1584	MF	-	-	-	-	-	-	-	-	CALR	-
1585	MF	-	-	-	-	-	-	-	-	CALR	-
1586	MF	-	-	-	-	-	-	-	-	JAK2V617F	Chr8,9
1587	MF	NFE2	12	54686469	_	delG	p.R271fs*31	0.12	frameshift (+1bp)	JAK2V617F	-
1588	MF	TET2	4	106164941	T	A	p.?	0.94	Splice	JAK2V617F	Chr9
1589	MF	DNMT3A	2	25463596	_	delG	p.Q696fs*9	0.41	frameshift (+1bp)	JAK2V617F	Chr9
1589	MF	NFE2	12	54686495	_	delCTCT	p.E261fs*3	0.02	frameshift (+1bp)	JAK2V617F	Chr9
1590	MF	-	-	-	-	-	-	-	-	JAK2V617F	Chr9
1591	MF	KRAS	12	25398242	T	A	p.N26I	0.54	Missense	JAK2V617F	Chr7,9
1591	MF	RUNX1	21	36252940	G	T	p.S141*	0.44	Nonsense	JAK2V617F	Chr7,9
1592	MF	-	-	-	-	-	-	-	-	JAK2V617F	Chr9
1593	MF	-	-	-	-	-	-	-	-	JAK2V617F	Chr9
1594	MF	ASXL1	20	31023717	C	T	p.R1068*	0.18	Nonsense	JAK2V617F	Chr9
1595	MF	-	-	-	-	-	-	-	-	JAK2V617F	Chr9,14
1596	MF	SRSF2	17	74732959	G	C	p.P95R	0.33	Missense	JAK2V617F	-
1597	MF	TET2	4	106155535	G	T	p.D146Y	0.21	Missense	JAK2V617F	-
1597	MF	MLL3	7	151945328	C	A	p.E731*	0.54	Nonsense	JAK2V617F	-
1597	MF	FLT3	13	28608260	T	A	p.Y599F	0.25	Missense	JAK2V617F	-
1598	MF	TET2	4	106156701	_	delAAAC	p.N535fs*6	0.28	frameshift (+1bp)	JAK2V617F	Chr9
1598	MF	IDH2	15	90631934	C	T	p.R140Q	0.16	Missense	JAK2V617F	Chr9
1599	MF	-	-	-	-	-	-	-	-	JAK2V617F	Chr9
1600	MF	TET2	4	106156767	_	insA	p.Q557fs*10	0.51	frameshift (+2bp)	JAK2V617F	Chr9

1601	MF	-	-	-	-	-	-	-	-	JAK2V617F	Chr9
1602	MF	-	-	-	-	-	-	-	-	JAK2V617F	Chr9
1603	MF	-	-	-	-	-	-	-	-	JAK2V617F	Chr9
1604	MF	-	-	-	-	-	-	-	-	JAK2V617F	-
1605	MF	TET2	4	106196684	_	delC	p.P1673fs*22	0.03	frameshift (+1bp)	JAK2V617F	Chr9
1606	MF	NFE2	12	54687087	_	insA	p.Y65fs*20	0.52	frameshift (+2bp)	JAK2V617F	Chr9
1607	MF	-	-	-	-	-	-	-	-	JAK2V617F	Chr9
1608	MF	PPM1D	17	58740602	_	insCAACAdelGTTG	p.S503fs*2	0.08	frameshift (+2bp)	JAK2V617F	Chr9
1609	MF	TET2	4	106196703	_	insT	p.Y1679fs*8	0.41	frameshift (+2bp)	JAK2V617F	Chr9
1609	MF	CBL	11	119169244	_	delACA	p.T811delT	0.7	Inframe	JAK2V617F	Chr9
1609	MF	NFE2	12	54686380	_	delCCGCTCCAGCTC	p.E297_R300delELER	0.5	Inframe	JAK2V617F	Chr9
1610	MF	TP53	17	7577551	C	A	p.G244C	0.81	Missense	JAK2V617F	Chr5,9,17
1610	MF	U2AF1	21	44514770	_	delCTCATA	p.Y158_E159delYE	0.36	Inframe	JAK2V617F	Chr5,9,17
1611	MF	-	-	-	-	-	-	-	-	JAK2V617F	Chr9
1612	MF	TET2	4	106156747	C	T	p.R550*	0.41	Nonsense	JAK2V617F	-
1612	MF	NFE2	12	54687044	_	delG	p.P79fs*32	0.28	frameshift (+1bp)	JAK2V617F	-
1613	MF	-	-	-	-	-	-	-	-	JAK2V617F	Chr9
1614	MF	NFE2	12	54686654	G	T	p.S209*	0.2	Nonsense	JAK2V617F	Chr9,20
1615	MF	DNMT3A	2	25463286	C	A	p.R736L	0.2	Missense	JAK2V617F	Chr9
1616	MF	DNMT3A	2	25457243	G	A	p.R882C	0.52	Missense	JAK2V617F	-
1617	MF	-	-	-	-	-	-	-	-	JAK2V617F	Chr9
1618	MF	ASXL1	20	31022839	T	A	p.L775*	0.21	Nonsense	JAK2V617F	Chr9
1619	MF	TET2	4	106156706	_	insT	p.K536fs*31	0.05	frameshift (+2bp)	JAK2V617F	Chr9
1620	MF	-	-	-	-	-	-	-	-	JAK2V617F	Chr9
1621	MF	-	-	-	-	-	-	-	-	JAK2V617F	Chr9
1622	MF	-	-	-	-	-	-	-	-	JAK2V617F	Chr9

1623	MF	-	-	-	-	-	-	-	-	JAK2V617F	Chr9
1624	MF	EZH2	7	148507443	G	C	p.L671V	0.13	Missense	JAK2V617F	Chr9
1624	MF	EZH2	7	148506167	A	C	p.Y731D	0.2	Missense	JAK2V617F	Chr9
1624	MF	ASXL1	20	31022403	_	delCACCA<13>GCGGC	p.E635fs*15	0.15	frameshift (+2bp)	JAK2V617F	Chr9
1624	MF	ASXL1	20	31022289	C	T	p.Q592*	0.17	Nonsense	JAK2V617F	Chr9
1625	MF	TET2	4	106164860	A	G	p.K1243R	0.48	Missense	JAK2V617F	Chr9
1625	MF	SH2B3	12	111885596	T	C	p.L458P	0.5	Missense	JAK2V617F	Chr9
1626	MF	ASXL1	20	31021630	_	delAGATCGTC	p.D544fs*4	0.07	frameshift (+2bp)	JAK2V617F	Chr9
1627	MF	NRAS	1	115256535	G	T	p.A59D	0.33	Missense	JAK2V617F	-
1627	MF	TET2	4	106164775	G	T	p.E1215*	0.4	Nonsense	JAK2V617F	-
1627	MF	TET2	4	106196705	C	T	p.Q1680*	0.45	Nonsense	JAK2V617F	-
1627	MF	PHF6	X	133551230	C	A	p.T289N	0.47	Missense	JAK2V617F	-
1628	MF	-	-	-	-	-	-	-	-	JAK2V617F	Chr9
1629	MF	TET2	4	106158248	C	A	p.S1050*	0.09	Nonsense	JAK2V617F	-
1629	MF	TP53	17	7577535	C	A	p.R249M	0.13	Missense	JAK2V617F	-
1629	MF	ASXL1	20	31022656	_	delC	p.R715fs*10	0.11	frameshift (+1bp)	JAK2V617F	-
1629	MF	ASXL1	20	31022287	_	insA	p.Y591fs*1	0.22	frameshift (+2bp)	JAK2V617F	-
1630	MF	-	-	-	-	-	-	-	-	JAK2V617F	-
1631	MF	TET2	4	106158455	T	G	p.L1119*	0.38	Nonsense	JAK2V617F	-
1632	MF	KRAS	12	25398285	C	T	p.G12S	0.21	Missense	JAK2V617F	Chr9
1632	MF	ASXL1	20	31022784	C	T	p.Q757*	0.41	Nonsense	JAK2V617F	Chr9
1633	MF	TET2	4	106180838	_	delG	p.C1289fs*74	0.78	frameshift (+1bp)	JAK2V617F	Chr9
1634	MF	PTPN11	12	112888210	G	C	p.E76Q	0.14	Missense	JAK2V617F	Chr9
1635	MF	-	-	-	-	-	-	-	-	JAK2V617F	Chr9
1636	MF	BCOR	X	39933295	_	delTC	p.D435fs*4	0.31	frameshift (+2bp)	JAK2V617F	-
1637	MPN_u	DNMT3A	2	25457242	C	T	p.R882H	0.38	Missense	JAK2V617F	-

1637	MPN_u	SF3B1	2	198267360	T	C	p.K666R	0.29	Missense	JAK2V617F	-
1637	MPN_u	SRSF2	17	74732960	G	C	p.P95A	0.39	Missense	JAK2V617F	-
1638	MPN_u	-	-	-	-	-	-	-	-	JAK2V617F	Chr9
1639	MPN_u	CUX1	7	101840530	_	delC	p.L614fs*25	0.34	frameshift (+1bp)	JAK2V617F	-
1639	MPN_u	U2AF1	21	44514777	T	G	p.Q157P	0.27	Missense	JAK2V617F	-
1640	MPN_u	-	-	-	-	-	-	-	-	CALR	-
1641	MPN_u	-	-	-	-	-	-	-	-	-	-
1642	MPN_u	-	-	-	-	-	-	-	-	-	-
1643	MPN_u	TET2	4	106157814	_	insA	p.M906fs*18	0.33	frameshift (+2bp)	JAK2V617F	-
1644	MPN_u	-	-	-	-	-	-	-	-	JAK2V617F	-
1645	MPN_u	-	-	-	-	-	-	-	-	JAK2V617F	-
1646	MPN_u	-	-	-	-	-	-	-	-	JAK2V617F	-
1647	MPN_u	-	-	-	-	-	-	-	-	JAK2V617F	-
1648	MPN_u	-	-	-	-	-	-	-	-	CALR	-
1649	MPN_u	-	-	-	-	-	-	-	-	JAK2V617F,CALR	-
1650	Atypical CML	GATA2	3	128205732	_	insA	p.F49fs*136	0.47	frameshift (+2bp)	-	-
1650	Atypical CML	SRSF2	17	74732959	G	A	p.P95L	0.53	Missense	-	-
1650	Atypical CML	ASXL1	20	31022403	_	delCACCA<13>GCGGC	p.E635fs*15	0.35	frameshift (+2bp)	-	-
1651	Other	-	-	-	-	-	-	-	-	-	-
1652	Other	-	-	-	-	-	-	-	-	-	-
1653	CMML	IDH2	15	90631934	C	T	p.R140Q	0.39	Missense	-	-
1653	CMML	SRSF2	17	74732959	G	A	p.P95L	0.35	Missense	-	-
1654	CMML	KRAS	12	25378652	T	G	p.N116H	0.44	Missense	-	-
1655	CMML	CBL	11	119148919	T	C	p.L380P	0.86	Missense	-	Chr11
1655	CMML	IDH2	15	90631934	C	T	p.R140Q	0.42	Missense	-	Chr11
1655	CMML	ASXL1	20	31022592	C	T	p.R693*	0.54	Nonsense	-	Chr11

1655	CMML	U2AF1	21	44514777	T	C	p.Q157R	0.38	Missense	-	Chr11
1656	CMML	KRAS	12	25378562	C	G	p.A146P	0.29	Missense	JAK2V617F	-
1656	CMML	U2AF1	21	44524456	G	A	p.S34F	0.34	Missense	JAK2V617F	-
1657	CMML	TET2	4	106190804	G	T	p.G1361V	0.41	Missense	-	Chr7
1657	CMML	TET2	4	106182954	_	delTCTT	p.L1332fs*30	0.43	frameshift (+1bp)	-	Chr7
1657	CMML	ASXL1	20	31021543	_	delTG	p.V515fs*13	0.38	frameshift (+2bp)	-	Chr7
1658	CMML	NRAS	1	115256532	C	A	p.G60V	0.31	Missense	-	-
1658	CMML	TET2	4	106190771	_	delA	p.E1350fs*13	0.34	frameshift (+1bp)	-	-
1658	CMML	TET2	4	106190827	T	C	p.S1369P	0.28	Missense	-	-
1658	CMML	SRSF2	17	74732959	G	A	p.P95L	0.35	Missense	-	-
1658	CMML	STAG2	X	123217380	C	T	p.R1012*	0.6	Nonsense	-	-
1659	Other	-	-	-	-	-	-	-	-	-	-
1660	Other	-	-	-	-	-	-	-	-	-	-
1661	Other	-	-	-	-	-	-	-	-	-	-
1662	IE	-	-	-	-	-	-	-	-	-	-
1663	Other	-	-	-	-	-	-	-	-	-	-
1664	IE	-	-	-	-	-	-	-	-	-	-
1665	IE	-	-	-	-	-	-	-	-	-	-
1666	IE	ASXL1	20	31022416	_	delGAGAG<13>CCATC	p.E635fs*15	0.17	frameshift (+2bp)	-	-
1667	Other	-	-	-	-	-	-	-	-	-	-
1668	Other	-	-	-	-	-	-	-	-	-	-
1669	IE	-	-	-	-	-	-	-	-	-	-
1670	Other	-	-	-	-	-	-	-	-	-	-
1671	Other	-	-	-	-	-	-	-	-	-	-
1672	MDS/MPD-SM	KIT	4	55599321	A	T	p.D816V	0.44	Missense	-	-
1672	MDS/MPD-SM	TET2	4	106196507	_	delC	p.M1615fs*1	0.48	frameshift (+1bp)	-	-

1672	MDS/MPD-SM	U2AF1	21	44524456	G	A	p.S34F	0.44	Missense	-	-
1673	MPN_u	EZH2	7	148523609	G	A	p.H282Y	0.56	Missense	JAK2V617F	-
1674	MPN_u	DNMT3A	2	25467498	G	C	p.Y526*	0.17	Nonsense	JAK2V617F	-
1674	MPN_u	TET2	4	106164897	_	delC	p.Y1255fs*1	0.04	frameshift (+1bp)	JAK2V617F	-
1675	Other	-	-	-	-	-	-	-	-	JAK2V617F	-
1676	RARST	SF3B1	2	198266834	T	C	p.K700E	0.15	Missense	JAK2V617F	Chr9
1677	RARST	SF3B1	2	198267360	T	C	p.K666R	0.47	Missense	JAK2V617F	-
1678	SM	TET2	4	106157644	_	delA	p.T849fs*24	0.12	frameshift (+1bp)	-	-
1679	SM	KIT	4	55599321	A	T	p.D816V	0.16	Missense	-	-
1680	SM	KIT	4	55599321	A	T	p.D816V	0.12	Missense	-	-
1681	Other	-	-	-	-	-	-	-	-	-	-
1682	SM	-	-	-	-	-	-	-	-	-	-
1683	Other	-	-	-	-	-	-	-	-	-	-
1684	Other	-	-	-	-	-	-	-	-	-	-
1685	PV	-	-	-	-	-	-	-	-	JAK2V617F	Chr9
1686	PV	NFE2	12	54686911	_	delG	p.L124fs*0	0.11	frameshift (+1bp)	JAK2V617F	Chr9
1687	PV	-	-	-	-	-	-	-	-	JAK2V617F	-
1688	PV	-	-	-	-	-	-	-	-	JAK2V617F	Chr9
1689	PV	TET2	4	106157002	C	T	p.Q635*	0.17	Nonsense	JAK2V617F	Chr9
1690	PV	-	-	-	-	-	-	-	-	JAK2V617F	-
1691	PV	TET2	4	106190797	C	T	p.R1359C	0.45	Missense	JAK2V617F	-
1691	PV	TET2	4	106156959	C	A	p.Y620*	0.42	Nonsense	JAK2V617F	-
1691	PV	ASXL1	20	31021472	C	T	p.Q491*	0.38	Nonsense	JAK2V617F	-
1692	PV	TET2	4	106196575	_	delC	p.L1637fs*58	0.06	frameshift (+1bp)	JAK2V617F	-
1693	PV	-	-	-	-	-	-	-	-	JAK2V617F	-
1694	PV	-	-	-	-	-	-	-	-	JAK2V617F	Chr9

1695	PV	-	-	-	-	-	-	-	-	JAK2V617F	-
1696	PV	-	-	-	-	-	-	-	-	JAK2V617F	Chr9
1697	PV	TET2	4	106180784	_	insG	p.C1271fs*29	0.33	frameshift (+2bp)	JAK2V617F	-
1698	PV	NF1	17	29533395	A	T	p.?	0.43	Splice	JAK2V617F	Chr9
1699	PV	-	-	-	-	-	-	-	-	JAK2V617F	Chr9
1700	PV	-	-	-	-	-	-	-	-	JAK2V617F	Chr9
1701	PV	TET2	4	106197347	C	T	p.P1894S	0.14	Missense	JAK2V617F	Chr9
1701	PV	EZH2	7	148512004	_	insA	p.?	0.39	Splice	JAK2V617F	Chr9
1701	PV	ASXL1	20	31022695	_	insCGGAA	p.E727fs*19	0.24	frameshift (+1bp)	JAK2V617F	Chr9
1702	PV	-	-	-	-	-	-	-	-	JAK2V617F	Chr9
1703	PV	-	-	-	-	-	-	-	-	JAK2V617F	-
1704	PV	-	-	-	-	-	-	-	-	JAK2V617F	Chr9
1705	PV	TET2	4	106155921	_	delC	p.N275fs*18	0.5	frameshift (+1bp)	JAK2V617F	-
1706	PV	-	-	-	-	-	-	-	-	JAK2V617F	-
1707	PV	-	-	-	-	-	-	-	-	JAK2V617F	Chr9
1708	PV	-	-	-	-	-	-	-	-	JAK2V617F	Chr18
1709	PV	-	-	-	-	-	-	-	-	JAK2V617F	-
1710	PV	-	-	-	-	-	-	-	-	JAK2V617F	Chr9
1711	PV	TET2	4	106164769	G	A	p.V1213M	0.2	Missense	JAK2V617F	-
1712	PV	TET2	4	106156339	C	T	p.Q414*	0.44	Nonsense	JAK2V617F	-
1713	PV	-	-	-	-	-	-	-	-	JAK2V617F	Chr9
1714	PV	-	-	-	-	-	-	-	-	JAK2V617F	-
1715	PV	-	-	-	-	-	-	-	-	JAK2V617F	Chr9
1716	PV	MLL3	7	151868399	C	A	p.G3135*	0.19	Nonsense	JAK2V617F	-
1716	PV	SRSF2	17	74732403	G	A	p.S169F	0.31	Missense	JAK2V617F	-
1717	PV	-	-	-	-	-	-	-	-	JAK2V617F	Chr9

1718	PV	TET2	4	106158483	_	insTTAT	p.D1129fs*2	0.45	frameshift (+2bp)	JAK2V617F	-
1719	PV	-	-	-	-	-	-	-	-	JAK2V617F	Chr9
1720	PV	STAG2	X	123195723	G	T	p.R546M	0.12	Missense	JAK2V617F	-
1721	PV	TET2	4	106157788	C	T	p.Q897*	0.39	Nonsense	JAK2V617F	Chr9
1721	PV	KRAS	12	25398220	A	T	p.D33E	0.12	Missense	JAK2V617F	Chr9
1721	PV	NFE2	12	54686495	_	delCT	p.S262fs*42	0.35	frameshift (+2bp)	JAK2V617F	Chr9
1722	PV	-	-	-	-	-	-	-	-	JAK2V617F	-
1723	PV	NFE2	12	54686495	_	delCTCT	p.E261fs*3	0.35	frameshift (+1bp)	JAK2V617F	Chr9
1724	PV	-	-	-	-	-	-	-	-	JAK2V617F	-
1725	PV	-	-	-	-	-	-	-	-	JAK2V617F	Chr9
1726	PV	-	-	-	-	-	-	-	-	JAK2V617F	Chr9
1727	PV	-	-	-	-	-	-	-	-	JAK2V617F	-
1728	PV	PTPN11	12	112888163	G	C	p.G60A	0.15	Missense	JAK2V617F	Chr9
1729	PV	-	-	-	-	-	-	-	-	JAK2V617F	-
1730	PV	-	-	-	-	-	-	-	-	JAK2V617F	Chr9
1731	PV	-	-	-	-	-	-	-	-	JAK2V617F	-
1732	PV	-	-	-	-	-	-	-	-	JAK2V617F	-
1733	PV	-	-	-	-	-	-	-	-	JAK2V617F	-
1734	PV	TET2	4	106156890	_	delCAAT	p.N598fs*2	0.38	frameshift (+1bp)	JAK2V617F	Chr9
1735	PV	DNMT3A	2	25463184	G	T	p.S770*	0.6	Nonsense	JAK2V617F	Chr9
1736	PV	TET2	4	106157560	C	T	p.Q821*	0.11	Nonsense	JAK2V617F	-
1737	PV	-	-	-	-	-	-	-	-	JAK2V617F	-
1738	PV	IDH1	2	209113112	C	T	p.R132H	0.22	Missense	JAK2V617F	Chr9,18
1739	PV	TET2	4	106180785	C	G	p.C1271W	0.48	Missense	JAK2V617F	Chr9
1740	PV	DNMT3A	2	25468887	A	C	p.?	0.33	Splice	JAK2V617F	Chr9
1741	PV	-	-	-	-	-	-	-	-	JAK2V617F	Chr9

1742	PV	-	-	-	-	-	-	-	-	JAK2V617F	-
1743	PV	-	-	-	-	-	-	-	-	JAK2V617F	-
1744	PV	-	-	-	-	-	-	-	-	JAK2V617F	Chr9
1745	PV	SF3B1	2	198267359	C	G	p.K666N	0.1	Missense	JAK2V617F	Chr9
1746	PV	TET2	4	106156747	C	T	p.R550*	0.46	Nonsense	JAK2V617F	Chr9
1747	PV	-	-	-	-	-	-	-	-	JAK2V617F	Chr9
1748	PV	-	-	-	-	-	-	-	-	JAK2V617F	Chr9
1749	PV	-	-	-	-	-	-	-	-	JAK2V617F	-
1750	PV	-	-	-	-	-	-	-	-	JAK2V617F	-
1751	PV	-	-	-	-	-	-	-	-	JAK2V617F	-
1752	PV	-	-	-	-	-	-	-	-	JAK2V617F	Chr9
1753	PV	-	-	-	-	-	-	-	-	JAK2V617F	Chr9
1754	PV	-	-	-	-	-	-	-	-	JAK2V617F	Chr9
1755	PV	-	-	-	-	-	-	-	-	JAK2V617F	-
1756	PV	-	-	-	-	-	-	-	-	JAK2V617F	-
1757	PV	EZH2	7	148508784	G	A	p.A627V	0.13	Missense	JAK2V617F	-
1758	PV	-	-	-	-	-	-	-	-	JAK2V617F	Chr9
1759	PV	-	-	-	-	-	-	-	-	JAK2V617F	-
1760	PV	-	-	-	-	-	-	-	-	JAK2V617F	-
1761	PV	PPM1D	17	58740529	C	A	p.C478*	0.47	Nonsense	JAK2V617F	Chr9
1762	PV	-	-	-	-	-	-	-	-	JAK2V617F	-
1763	PV	ZRSR2	X	15838446	G	T	p.?	0.15	Splice	JAK2V617F	-
1764	PV	-	-	-	-	-	-	-	-	JAK2V617F	-
1765	PV	-	-	-	-	-	-	-	-	JAK2V617F	Chr9
1766	PV	-	-	-	-	-	-	-	-	JAK2V617F	-
1767	PV	DNMT3A	2	25457161	A	C	p.F909C	0.33	Missense	JAK2V617F	Chr9

1768	PV	-	-	-	-	-	-	-	-	JAK2V617F	-
1769	PV	-	-	-	-	-	-	-	-	JAK2V617F	-
1770	PV	-	-	-	-	-	-	-	-	JAK2V617F	-
1771	PV	-	-	-	-	-	-	-	-	JAK2V617F	-
1772	PV	TET2	4	106196700	_	insT	p.Y1679fs*8	0.3	frameshift (+2bp)	JAK2V617F	Chr9
1772	PV	CBL	11	119169244	_	delACA	p.T811delT	0.5	Inframe	JAK2V617F	Chr9
1772	PV	NFE2	12	54686380	_	delCCGCTCCAGCTC	p.E297_R300delELER	0.47	Inframe	JAK2V617F	Chr9
1773	PV	-	-	-	-	-	-	-	-	JAK2V617F	-
1774	PV	TET2	4	106157240	C	G	p.S714*	0.29	Nonsense	JAK2V617F	Chr9
1774	PV	EZH2	7	148506443	C	T	p.R690H	0.29	Missense	JAK2V617F	Chr9
1775	PV	-	-	-	-	-	-	-	-	JAK2V617F	Chr9
1776	PV	-	-	-	-	-	-	-	-	JAK2V617F	Chr9
1777	PV	-	-	-	-	-	-	-	-	JAK2V617F	-
1778	PV	ASXL1	20	31022834	_	delGAGAT	p.E773fs*12	0.33	frameshift (+2bp)	JAK2V617F	Chr9
1779	PV	-	-	-	-	-	-	-	-	JAK2V617F	-
1780	PV	TET2	4	106180793	_	delAG	p.Q1274fs*25	0.38	frameshift (+2bp)	JAK2V617F	Chr9
1781	PV	-	-	-	-	-	-	-	-	JAK2V617F	-
1782	PV	-	-	-	-	-	-	-	-	JAK2V617F	-
1783	PV	-	-	-	-	-	-	-	-	JAK2V617F	-
1784	PV	-	-	-	-	-	-	-	-	JAK2V617F	Chr9
1785	PV	TET2	4	106193931	C	T	p.R1465*	0.18	Nonsense	JAK2V617F	-
1785	PV	ASXL1	20	31022880	_	delT	p.C789fs*29	0.54	frameshift (+1bp)	JAK2V617F	-
1786	PV	-	-	-	-	-	-	-	-	JAK2V617F	-
1787	PV	-	-	-	-	-	-	-	-	JAK2V617F	Chr9
1788	PV	TET2	4	106164914	G	T	p.R1261L	0.23	Missense	JAK2V617F	-
1789	PV	-	-	-	-	-	-	-	-	JAK2V617F	-

1790	PV	NFE2	12	54686495	_	delCTCT	p.E261fs*3	0.45	frameshift (+1bp)	JAK2V617F	Chr9
1791	PV	TET2	4	106197551	C	T	p.P1962S	0.15	Missense	JAK2V617F	Chr9
1792	PV	-	-	-	-	-	-	-	-	JAK2V617F	Chr9,20
1793	PV	TET2	4	106158124	C	T	p.Q1009*	0.28	Nonsense	JAK2V617F	-
1794	PV	TET2	4	106164940	G	A	p.?	0.53	Splice	JAK2V617F	-
1794	PV	TET2	4	106180783	T	G	p.C1271G	0.32	Missense	JAK2V617F	-
1795	PV	-	-	-	-	-	-	-	-	JAK2V617F	Chr9
1796	PV	-	-	-	-	-	-	-	-	JAK2V617F	Chr9
1797	PV	-	-	-	-	-	-	-	-	JAK2V617F	Chr9
1798	PV	TET2	4	106197317	A	G	p.T1884A	0.24	Missense	JAK2V617F	-
1799	PV	GNB1	1	1747229	T	C	p.K57E	0.19	Missense	JAK2V617F	-
1800	PV	ASXL1	20	31022288	C	A	p.Y591*	0.62	Nonsense	JAK2V617F	-
1801	PV	-	-	-	-	-	-	-	-	JAK2V617F	-
1802	PV	-	-	-	-	-	-	-	-	JAK2V617F	-
1803	PV	-	-	-	-	-	-	-	-	JAK2V617F	Chr9
1804	PV	-	-	-	-	-	-	-	-	JAK2V617F	-
1805	PV	TET2	4	106156359	_	delA	p.E421fs*6	0.38	frameshift (+1bp)	JAK2V617F	Chr9
1805	PV	TET2	4	106193778	C	T	p.Q1414*	0.47	Nonsense	JAK2V617F	Chr9
1806	PV	-	-	-	-	-	-	-	-	JAK2V617F	-
1807	PV	-	-	-	-	-	-	-	-	JAK2V617F	Chr9
1808	PV	-	-	-	-	-	-	-	-	JAK2V617F	-
1809	PV	NF1	17	29527531	_	delTG	p.C328fs*1	0.06	frameshift (+2bp)	JAK2V617F	-
1810	PV	ASXL1	20	31023273	_	insA	p.P920fs*4	1	frameshift (+2bp)	JAK2V617F	Chr9
1811	PV	-	-	-	-	-	-	-	-	JAK2V617F	Chr9
1812	PV	-	-	-	-	-	-	-	-	JAK2V617F	-
1813	PV	-	-	-	-	-	-	-	-	JAK2V617F	Chr9

1814	PV	-	-	-	-	-	-	-	-	JAK2V617F	Chr9
1815	PV	-	-	-	-	-	-	-	-	JAK2V617F	Chr9
1816	PV	-	-	-	-	-	-	-	-	JAK2V617F	Chr9
1817	PV	-	-	-	-	-	-	-	-	JAK2V617F	Chr9
1818	PV	DNMT3A	2	25463184	G	T	p.S770*	0.26	Nonsense	JAK2V617F	Chr9
1819	PV	-	-	-	-	-	-	-	-	JAK2V617F	-
1820	PV	-	-	-	-	-	-	-	-	JAK2V617F	Chr9
1821	PV	-	-	-	-	-	-	-	-	JAK2V617F	Chr9
1822	PV	-	-	-	-	-	-	-	-	JAK2V617F	-
1823	PV	-	-	-	-	-	-	-	-	JAK2V617F	Chr9
1824	PV	-	-	-	-	-	-	-	-	JAK2V617F	Chr9
1825	PV	TET2	4	106196324	_	delCAGA	p.T1554fs*16	0.35	frameshift (+1bp)	JAK2V617F	Chr9,14
1825	PV	BCOR	X	39932209	_	delAC	p.V797fs*19	0.32	frameshift (+2bp)	JAK2V617F	Chr9,14
1826	PV	STAG2	X	123179222	G	T	p.?	0.24	Splice	JAK2V617F	-
1827	PV	-	-	-	-	-	-	-	-	JAK2V617F	-
1828	PV	-	-	-	-	-	-	-	-	JAK2V617F	-
1829	PV	NFE2	12	54686429	C	T	p.R284H	0.52	Missense	JAK2V617F	Chr9
1830	PV	-	-	-	-	-	-	-	-	JAK2V617F	Chr9
1831	PV	-	-	-	-	-	-	-	-	JAK2V617F	Chr9
1832	PV	STAG2	X	123179092	G	A	p.V181M	0.48	Missense	JAK2V617F	Chr9
1833	PV	-	-	-	-	-	-	-	-	JAK2V617F	-
1834	PV	NFE2	12	54686495	_	delCTCT	p.E261fs*3	0.26	frameshift (+1bp)	JAK2V617F	Chr9
1835	PV	-	-	-	-	-	-	-	-	JAK2V617F	-
1836	PV	TET2	4	106190767	_	insA	p.I1349fs*2	0.12	frameshift (+2bp)	JAK2V617F	Chr9
1836	PV	TET2	4	106164913	C	A	p.R1261S	0.28	Missense	JAK2V617F	Chr9
1837	PV	-	-	-	-	-	-	-	-	JAK2V617F	-

1838	PV	-	-	-	-	-	-	-	-	JAK2V617F	-
1839	PV	-	-	-	-	-	-	-	-	JAK2V617F	Chr9
1840	PV	-	-	-	-	-	-	-	-	JAK2V617F	Chr9
1841	PV	-	-	-	-	-	-	-	-	JAK2V617F	-
1842	PV	-	-	-	-	-	-	-	-	JAK2V617F	Chr9
1843	PV	-	-	-	-	-	-	-	-	JAK2V617F	-
1844	PV	TET2	4	106156237	_	delTACTT<9>CTCAG	p.Y380fs*41	0.35	frameshift (+1bp)	JAK2V617F	Chr9
1845	PV	-	-	-	-	-	-	-	-	JAK2V617F	-
1846	PV	KRAS	12	25380285	G	C	p.T58R	0.29	Missense	JAK2V617F	Chr9
1847	PV	JAK2	9	5089783	A	G	p.D894G	0.27	Missense	JAK2V617F	-
1848	PV	-	-	-	-	-	-	-	-	JAK2V617F	Chr9
1849	PV	-	-	-	-	-	-	-	-	JAK2V617F	Chr9
1850	PV	STAG2	X	123191825	G	T	p.E472*	0.3	Nonsense	JAK2V617F	Chr9
1851	PV	NFE2	12	54686577	_	insA	p.P236fs*14	0.45	frameshift (+2bp)	JAK2V617F	Chr9
1852	PV	-	-	-	-	-	-	-	-	JAK2V617F	Chr9
1853	PV	-	-	-	-	-	-	-	-	JAK2V617F	Chr9
1854	PV	-	-	-	-	-	-	-	-	JAK2V617F	Chr9
1855	PV	-	-	-	-	-	-	-	-	JAK2V617F	-
1856	PV	-	-	-	-	-	-	-	-	JAK2V617F	-
1857	PV	-	-	-	-	-	-	-	-	JAK2V617F	Chr9
1858	PV	-	-	-	-	-	-	-	-	JAK2V617F	Chr9
1859	PV	-	-	-	-	-	-	-	-	JAK2V617F	Chr9
1860	PV	-	-	-	-	-	-	-	-	JAK2V617F	Chr9
1861	PV	TET2	4	106156043	_	delC	p.Q317fs*30	0.01	frameshift (+1bp)	JAK2V617F	-
1862	PV	TET2	4	106193853	A	T	p.K1439*	0.12	Nonsense	JAK2V617F	Chr9
1863	PV	TET2	4	106180817	G	A	p.G1282D	0.53	Missense	JAK2V617F	-

1863	PV	TET2	4	106164897	_	insA	p.Y1255fs*1	0.42	frameshift (+2bp)	JAK2V617F	-
1864	PV	-	-	-	-	-	-	-	-	JAK2V617F	-
1865	PV	-	-	-	-	-	-	-	-	JAK2V617F	Chr9
1866	PV	TET2	4	106164074	_	insT	p.A1196fs*2	0.09	frameshift (+2bp)	JAK2V617F	Chr9
1867	PV	-	-	-	-	-	-	-	-	JAK2V617F	Chr9
1868	PV	-	-	-	-	-	-	-	-	JAK2V617F	-
1869	PV	-	-	-	-	-	-	-	-	JAK2exon12	-
1870	PV	-	-	-	-	-	-	-	-	JAK2V617F	-
1871	PV	-	-	-	-	-	-	-	-	JAK2V617F	-
1872	PV	TET2	4	106196561	C	T	p.Q1632*	0.47	Nonsense	JAK2V617F	Chr7,9
1872	PV	PPM1D	17	58740749	C	T	p.R552*	0.06	Nonsense	JAK2V617F	Chr7,9
1873	PV	-	-	-	-	-	-	-	-	JAK2V617F	Chr9
1874	PV	TET2	4	106157638	C	T	p.Q847*	0.37	Nonsense	JAK2V617F	Chr9
1875	PV	-	-	-	-	-	-	-	-	JAK2V617F	Chr9
1876	PV	-	-	-	-	-	-	-	-	JAK2V617F	Chr9
1877	PV	DNMT3A	2	25463322	_	delGA	p.?	0.38	Splice	JAK2V617F	Chr9,14
1878	PV	SF3B1	2	198265519	C	T	p.G880R	0.22	Missense	JAK2V617F	Chr9
1878	PV	FLT3	13	28622545	C	T	p.D358N	0.18	Missense	JAK2V617F	Chr9
1879	PV	TET2	4	106157986	_	delC	p.Q963fs*44	0.25	frameshift (+1bp)	JAK2V617F	-
1880	PV	-	-	-	-	-	-	-	-	JAK2V617F	-
1881	PV	-	-	-	-	-	-	-	-	JAK2V617F	-
1882	PV	DNMT3A	2	25464534	T	A	p.Y660F	0.12	Missense	JAK2V617F	Chr9
1882	PV	TET2	4	106155938	_	insT	p.N281fs*0	0.26	frameshift (+2bp)	JAK2V617F	Chr9
1883	PV	-	-	-	-	-	-	-	-	JAK2V617F	Chr9
1884	PV	-	-	-	-	-	-	-	-	JAK2V617F	-
1885	PV	-	-	-	-	-	-	-	-	JAK2V617F	-

1886	PV	TET2	4	106155862	C	T	p.Q255*	0.39	Nonsense	JAK2V617F	Chr9
1887	PV	-	-	-	-	-	-	-	-	JAK2V617F	Chr9
1888	PV	-	-	-	-	-	-	-	-	JAK2V617F	-
1889	PV	TET2	4	106196993	_	delT	p.S1776fs*44	0.48	frameshift (+1bp)	JAK2V617F	Chr9
1890	PV	-	-	-	-	-	-	-	-	JAK2V617F	Chr9
1891	PV	NFE2	12	54686380	_	delCCGCTCCAGCTC	p.E297_R300delELER	0.07	Inframe	JAK2V617F	-
1892	PV	-	-	-	-	-	-	-	-	JAK2V617F	-
1893	PV	-	-	-	-	-	-	-	-	JAK2V617F	-
1894	PV	-	-	-	-	-	-	-	-	JAK2V617F	Chr9
1895	PV	STAG2	X	123227935	G	A	p.E1216K	0.13	Missense	JAK2V617F	-
1896	PV	CBL	11	119146791	C	G	p.I318M	0.25	Missense	JAK2V617F	Chr9
1897	PV	-	-	-	-	-	-	-	-	JAK2V617F	Chr9
1898	PV	DNMT3A	2	25463568	A	G	p.I705T	0.19	Missense	JAK2V617F	-
1899	PV	-	-	-	-	-	-	-	-	JAK2V617F	Chr9
1900	PV	-	-	-	-	-	-	-	-	JAK2V617F	Chr9
1901	PV	-	-	-	-	-	-	-	-	JAK2V617F	Chr9
1902	PV	-	-	-	-	-	-	-	-	JAK2V617F	Chr9
1903	PV	NFE2	12	54686495	_	delCTCT	p.E261fs*3	0.09	frameshift (+1bp)	JAK2V617F	Chr9
1904	PV	-	-	-	-	-	-	-	-	JAK2V617F	-
1905	PV	TET2	4	106158509	G	A	p.?	0.07	Splice	JAK2V617F	-
1905	PV	TET2	4	106157653	G	T	p.E852*	0.33	Nonsense	JAK2V617F	-
1906	PV	EZH2	7	148508779	A	G	p.W629R	0.11	Missense	JAK2V617F	Chr11
1907	PV	DNMT3A	2	25469055	_	insTTGACCT	p.E469fs*6	0.08	frameshift (+2bp)	JAK2V617F	-
1908	PV	-	-	-	-	-	-	-	-	JAK2V617F	-
1909	PV	-	-	-	-	-	-	-	-	JAK2V617F	Chr9
1910	PV	ASXL1	20	31021211	C	T	p.R404*	0.45	Nonsense	JAK2V617F	Chr9

1911	PV	-	-	-	-	-	-	-	-	JAK2V617F	-
1912	PV	-	-	-	-	-	-	-	-	JAK2V617F	Chr9
1913	PV	-	-	-	-	-	-	-	-	JAK2V617F	Chr9
1914	PV	-	-	-	-	-	-	-	-	JAK2V617F	-
1915	PV	TET2	4	106180838	_	delG	p.C1289fs*74	0.44	frameshift (+1bp)	JAK2V617F	Chr9
1916	PV	SF3B1	2	198266611	C	T	p.G742D	0.28	Missense	JAK2V617F	Chr9
1917	PV	DNMT3A	2	25470498	G	A	p.R326C	0.14	Missense	JAK2V617F	Chr9
1917	PV	BCOR	X	39930312	_	delCAGT	p.D1050fs*4	0.18	frameshift (+1bp)	JAK2V617F	Chr9
1918	PV	-	-	-	-	-	-	-	-	JAK2V617F	-
1919	PV	-	-	-	-	-	-	-	-	JAK2V617F	-
1920	PV	-	-	-	-	-	-	-	-	JAK2V617F	Chr9
1921	PV	-	-	-	-	-	-	-	-	JAK2V617F	-
1922	PV	TET2	4	106155749	_	delC	p.V218fs*32	0.07	frameshift (+1bp)	JAK2V617F	-
1923	PV	-	-	-	-	-	-	-	-	JAK2V617F	-
1924	PV	NFE2	12	54687089	_	insA	p.Y65fs*20	0.51	frameshift (+2bp)	JAK2V617F	Chr9
1924	PV	ASXL1	20	31023571	_	insG	p.E1020fs*14	0.57	frameshift (+2bp)	JAK2V617F	Chr9
1925	PV	TET2	4	106157067	_	delCTCA	p.S657fs*42	0.41	frameshift (+1bp)	JAK2V617F	-
1925	PV	PPM1D	17	58740543	_	insAG	p.T483fs*3	0.39	frameshift (+1bp)	JAK2V617F	-
1926	PV	-	-	-	-	-	-	-	-	JAK2exon12	Chr9
1927	PV	-	-	-	-	-	-	-	-	JAK2V617F	-
1928	PV	-	-	-	-	-	-	-	-	JAK2exon12	-
1929	PV	TET2	4	106157297	_	delC	p.Q734fs*17	0.07	frameshift (+1bp)	JAK2V617F	Chr9
1930	PV	DNMT3A	2	25457163	A	T	p.Y908*	0.34	Nonsense	JAK2exon12	-
1931	PV	-	-	-	-	-	-	-	-	JAK2V617F	Chr9
1932	PV	-	-	-	-	-	-	-	-	JAK2V617F	-
1933	PV	NFE2	12	54686575	_	delA	p.P236fs*12	0.55	frameshift (+1bp)	JAK2V617F	Chr9

1934	PV	TET2	4	106196702	_	insT	p.Y1679fs*8	0.42	frameshift (+2bp)	JAK2V617F	-
1935	PV	PPM1D	17	58740714	_	delA	p.E540fs*7	0.07	frameshift (+1bp)	JAK2V617F	Chr9
1936	PV	STAG2	X	123185054	T	A	p.F367L	0.14	Missense	JAK2V617F	Chr1q,9
1937	PV	TET2	4	106164940	G	T	p.?	0.47	Splice	JAK2V617F	Chr9
1937	PV	JAK2	9	5073764	G	C	p.V615L	0.89	Missense	JAK2V617F	Chr9
1937	PV	PPM1D	17	58740444	T	A	p.L450*	0.16	Nonsense	JAK2V617F	Chr9
1937	PV	TP53	17	7578268	A	C	p.L194R	0.16	Missense	JAK2V617F	Chr9
1937	PV	GNAS	20	57484421	G	A	p. R201H	0.17	Missense	JAK2V617F	Chr9
1938	PV	-	-	-	-	-	-	-	-	JAK2V617F	-
1939	PV	-	-	-	-	-	-	-	-	JAK2V617F	Chr9
1940	PV	-	-	-	-	-	-	-	-	JAK2V617F	-
1941	PV	-	-	-	-	-	-	-	-	JAK2V617F	Chr9
1942	PV	IDH1	2	209113112	C	T	p.R132H	0.17	Missense	JAK2V617F	Chr9
1943	PV	DNMT3A	2	25457176	G	A	p.P904L	0.23	Missense	JAK2V617F	Chr9
1943	PV	TET2	4	106190802	_	delG	p.G1361fs*2	0.17	frameshift (+1bp)	JAK2V617F	Chr9
1944	PV	TET2	4	106158433	_	delCTAGATACTC	p.D1113fs*1	0.45	frameshift (+1bp)	JAK2V617F	Chr9
1944	PV	TET2	4	106158425	_	delCC	p.S1109fs*1	0.54	frameshift (+2bp)	JAK2V617F	Chr9
1945	PV	-	-	-	-	-	-	-	-	JAK2V617F	-
1946	PV	NFE2	12	54686314	_	insA	p.R323fs*10	0.31	frameshift (+2bp)	JAK2V617F	Chr9
1947	PV	-	-	-	-	-	-	-	-	JAK2V617F	-
1948	PV	-	-	-	-	-	-	-	-	JAK2V617F	-
1949	PV	-	-	-	-	-	-	-	-	JAK2V617F	Chr8,9
1950	PV	-	-	-	-	-	-	-	-	JAK2V617F	Chr9
1951	PV	-	-	-	-	-	-	-	-	JAK2V617F	-
1952	PV	TET2	4	106197373	C	T	p.Y1902Y	0.39	Missense	JAK2V617F	Chr9
1953	PV	-	-	-	-	-	-	-	-	JAK2V617F	-

1954	PV	-	-	-	-	-	-	-	-	JAK2V617F	Chr9
1955	PV	SF3B1	2	198266822	T	A	p.I704F	0.09	Missense	JAK2V617F	Chr11
1956	PV	-	-	-	-	-	-	-	-	JAK2V617F	Chr9
1957	PV	-	-	-	-	-	-	-	-	JAK2V617F	-
1958	PV	-	-	-	-	-	-	-	-	JAK2exon12	-
1959	PV	-	-	-	-	-	-	-	-	JAK2V617F	Chr9
1960	PV	-	-	-	-	-	-	-	-	JAK2V617F	Chr5,9
1961	PV	-	-	-	-	-	-	-	-	JAK2V617F	-
1962	PV	-	-	-	-	-	-	-	-	JAK2V617F	-
1963	PV	-	-	-	-	-	-	-	-	JAK2V617F	-
1964	PV	-	-	-	-	-	-	-	-	JAK2V617F	-
1965	PV	TET2	4	106158455	T	G	p.L1119*	0.34	Nonsense	JAK2V617F	-
1966	PV	NFE2	12	54686495	_	delCTCT	p.E261fs*3	0.09	frameshift (+1bp)	JAK2V617F	Chr9
1966	PV	PTPN11	12	112924336	G	A	p.V428M	0.47	Missense	JAK2V617F	Chr9
1967	PV	-	-	-	-	-	-	-	-	JAK2V617F	Chr8,9
1968	PV	-	-	-	-	-	-	-	-	JAK2V617F	-
1969	PV	-	-	-	-	-	-	-	-	JAK2exon12	Chr9
1970	PV	-	-	-	-	-	-	-	-	JAK2V617F	Chr9
1971	PV	RUNX1	21	36252872	C	T	p.V164I	0.55	Missense	JAK2V617F	Chr9
1972	PV	KIT	4	55599285	G	A	p.R804Q	0.51	Missense	JAK2V617F	-
1973	PV	TET2	4	106158301	_	delC	p.Q1068fs*14	0.04	frameshift (+1bp)	JAK2V617F	-
1974	PV	-	-	-	-	-	-	-	-	JAK2V617F	Chr9
1975	PV	MLL3	7	151945354	_	delCCTTGTAG	p.L720fs*9	0.37	frameshift (+2bp)	JAK2V617F	Chr9
1976	PV	-	-	-	-	-	-	-	-	JAK2V617F	Chr9
1977	PV	TET2	4	106197115	_	insA	p.H1817fs*5	0.15	frameshift (+2bp)	JAK2V617F	-
1977	PV	ASXL1	20	31021540	_	delAACTGTGGATC	p.T514fs*11	0.39	frameshift (+2bp)	JAK2V617F	-

1978	PV	-	-	-	-	-	-	-	-	JAK2V617F	Chr9
1979	PV	BCOR	X	39933575	G	A	p.R342*	0.8	Nonsense	JAK2V617F	Chr9
1980	PV	-	-	-	-	-	-	-	-	JAK2V617F	Chr9
1981	PV	NF1	17	29588751	C	T	p.R1534*	0.06	Nonsense	JAK2V617F	Chr9
1982	PV	-	-	-	-	-	-	-	-	JAK2V617F	Chr9
1983	PV	NFE2	12	54686495	_	delCTCT	p.E261fs*3	0.02	frameshift (+1bp)	JAK2V617F	Chr9
1983	PV	PTPN11	12	112924336	G	A	p.V428M	0.23	Missense	JAK2V617F	Chr9
1984	PV	TET2	4	106155594	_	delTT	p.S167fs*3	0.89	frameshift (+2bp)	JAK2V617F	Chr9
1985	PV	NFE2	12	54686380	_	delCCGCTCCAGCTC	p.E297_R300delELER	0.33	Inframe	JAK2V617F	Chr9
1986	MPN_u	-	-	-	-	-	-	-	-	CALR	-
1987	PV	SH2B3	12	111884838	G	T	p.?	0.07	Splice	JAK2V617F	Chr1q,9
1988	PV	EZH2	7	148523609	G	A	p.H282Y	0.47	Missense	JAK2V617F	Chr7,9
1989	PV	-	-	-	-	-	-	-	-	JAK2V617F	Chr1q,7,9
1990	PV	DNMT3A	2	25467498	G	C	p.Y526*	0.06	Nonsense	JAK2V617F	-
1991	PV	-	-	-	-	-	-	-	-	JAK2V617F	Chr8,9
1992	PV	-	-	-	-	-	-	-	-	JAK2V617F	-
1993	PV	-	-	-	-	-	-	-	-	JAK2V617F	Chr9
1994	PV	-	-	-	-	-	-	-	-	JAK2V617F	-
1995	PV	-	-	-	-	-	-	-	-	JAK2V617F	-
1996	PV	-	-	-	-	-	-	-	-	JAK2V617F	-
1997	PV	-	-	-	-	-	-	-	-	JAK2V617F	-
1998	PV	GNAS	20	57484421	G	A	p. R201H	0.48	Missense	JAK2V617F	Chr9
1999	PV	-	-	-	-	-	-	-	-	JAK2V617F	Chr9,18
2000	PV	EZH2	7	148515170	G	A	p.R347W	0.53	Missense	JAK2V617F	Chr9
2000	PV	NFE2	12	54687044	_	delG	p.P79fs*32	0.37	frameshift (+1bp)	JAK2V617F	Chr9
2001	PV	-	-	-	-	-	-	-	-	JAK2V617F	Chr9

2002	PV	-	-	-	-	-	-	-	-	JAK2V617F	Chr9
2003	PV	-	-	-	-	-	-	-	-	JAK2V617F	Chr9,17
2004	PV	ASXL1	20	31023501	_	delG	p.G996fs*28	0.16	frameshift (+1bp)	JAK2V617F	-
2005	PV	-	-	-	-	-	-	-	-	JAK2V617F	-
2006	PV	-	-	-	-	-	-	-	-	JAK2V617F	-
2007	PV	DNMT3A	2	25458575	C	T	p.?	0.07	Splice	JAK2V617F	-
2008	PV	-	-	-	-	-	-	-	-	JAK2V617F	-
2009	PV	-	-	-	-	-	-	-	-	JAK2V617F	Chr9
2010	PV	TP53	17	7579591	_	insA	p.S33fs*10	0.08	frameshift (+2bp)	JAK2V617F	-
2011	PV	-	-	-	-	-	-	-	-	JAK2V617F	-
2012	PV	TET2	4	106156043	_	delC	p.Q317fs*30	0.39	frameshift (+1bp)	JAK2V617F	Chr7
2012	PV	TET2	4	106155749	_	delC	p.V218fs*32	0.03	frameshift (+1bp)	JAK2V617F	Chr7
2013	PV	-	-	-	-	-	-	-	-	JAK2V617F	-
2014	PV	-	-	-	-	-	-	-	-	JAK2V617F	-
2015	PV	-	-	-	-	-	-	-	-	JAK2V617F	Chr9
2016	PV	-	-	-	-	-	-	-	-	JAK2V617F	Chr9
2017	PV	-	-	-	-	-	-	-	-	JAK2V617F	Chr9
2018	PV	-	-	-	-	-	-	-	-	JAK2V617F	-
2019	PV	-	-	-	-	-	-	-	-	JAK2V617F	Chr9
2020	PV	-	-	-	-	-	-	-	-	JAK2V617F	Chr9
2021	PV	SF3B1	2	198266584	C	A	p.G751V	0.06	Missense	JAK2V617F	Chr8,9
2022	PV	-	-	-	-	-	-	-	-	JAK2V617F	-
2023	PV	-	-	-	-	-	-	-	-	JAK2V617F	-
2024	PV	-	-	-	-	-	-	-	-	JAK2V617F	Chr9,11
2025	PV	-	-	-	-	-	-	-	-	JAK2V617F	-
2026	PV	-	-	-	-	-	-	-	-	JAK2V617F	-

2027	PV	-	-	-	-	-	-	-	-	JAK2V617F	-
2028	PV	DNMT3A	2	25469525	G	A	p.Q415*	0.38	Nonsense	JAK2V617F	Chr9
2029	PV	-	-	-	-	-	-	-	-	JAK2V617F	-
2030	PV	-	-	-	-	-	-	-	-	JAK2V617F	Chr9
2031	PV	-	-	-	-	-	-	-	-	JAK2V617F	Chr9
2032	PV	TET2	4	106196562	_	delA	p.Q1632fs*63	0.31	frameshift (+1bp)	JAK2V617F	Chr9
2033	PV	-	-	-	-	-	-	-	-	JAK2V617F	Chr9
2034	PV	-	-	-	-	-	-	-	-	JAK2V617F	Chr9
2035	PV	PHF6	X	133559280	G	A	p.E340K	0.11	Missense	JAK2V617F	Chr9
2036	PV	-	-	-	-	-	-	-	-	JAK2V617F	-
2037	PV	GNB1	1	1747229	T	C	p.K57E	0.48	Missense	JAK2V617F	-
2038	PV	TP53	17	7577120	C	T	p.R273H	0.55	Missense	-	Chr5,7
2039	PV	-	-	-	-	-	-	-	-	JAK2V617F	Chr9
2040	PV	-	-	-	-	-	-	-	-	JAK2V617F	-
2041	PV	-	-	-	-	-	-	-	-	JAK2V617F	Chr9

Appendix 4: Accuracy of the multivariate multistate model in patients with MF

Accuracy of the multivariate multistate model was assessed using several statistical parameters. Results of Brier score ("Brier") and concordance ("Uno C") statistics are shown using the multivariate multistate model ("full model") applied to the training cohort using leave-one-out cross-validation ("Training-Xval"), as well as to the external cohort ("External"). Also shown is the concordance statistic obtained using currently available prognostic schemas (High Molecular Risk, "HMR"; international prognostic scoring system ("IPSS"); Dynamic IPSS ("DIPSS"). Outcomes assessed using overall survival ("OS") and AML transformation ("AML").

Diagnosis	Cohort	Model	Outcome	Time (yrs)	Brier	Absolute Prediction Error	Stderr	Uno C	Uncertainty
MF	Training-Xval	Full model	OS	5	0.15	0.31	0.03	0.85	0.25
MF	Training-Xval	Full model	OS	10	0.13	0.29	0.03	0.85	0.14
MF	Training-Xval	Full model	OS	15	0.09	0.23	0.03	0.86	0.06
MF	Training-Xval	HMR	OS	-	-	-	0.02	0.65	-
MF	Training-Xval	Full model	AMLT	5	0.08	0.14	0.06	0.85	0.25
MF	Training-Xval	Full model	AMLT	10	0.09	0.16	0.06	0.84	0.17
MF	Training-Xval	Full model	AMLT	15	0.08	0.17	0.06	0.76	0.06
MF	Training-Xval	HMR	AMLT	-	-	-	0.04	0.64	-
MF	External	Limited genomic data	OS	5	0.12	0.28	0.03	0.81	0.24
MF	External	Limited genomic data	OS	10	0.1	0.24	0.03	0.78	0.12
MF	External	Limited genomic data	OS	15	0.06	0.18	0.03	0.7	0.04
MF	External	HMR	OS	-	-	-	0.03	-	-
MF	External	IPSS	OS	-	-	-	0.03	0.82	-
MF	External	DIPSS	OS	-	-	-	0.03	0.68	-
MF	External	Limited genomic data	AMLT	5	0.08	0.16	0.06	0.85	0.25
MF	External	Limited genomic data	AMLT	10	0.08	0.18	0.06	0.84	0.14
MF	External	Limited genomic data	AMLT	15	0.08	0.2	0.06	0.81	0.07
MF	External	HMR	AMLT	-	-	-	0.05	-	-
MF	External	IPSS	AMLT	-	-	-	0.05	0.74	-
MF	External	DIPSS	AMLT	-	-	-	0.05	0.76	-

Appendix 5

Assessment of performance of multistate model in chronic phase (CP) patients

Results of Brier score ("Brier") and concordance ("Uno C") statistics are shown using the multivariate multistate model ("full model") applied to the training cohort using leave-one-out cross-validation ("Training (Xval)"), as well as to the external cohort ("External"). These are compared to the concordance obtained using the International Prognostic Score in ET (IPSET) for patients with ET. Outcomes are overall and event free survival (OS and EFS), and transformation to myelofibrosis (MFT) or acute myeloid leukemia (AMLT). Also shown is the concordance statistic obtained using currently available prognostic schemas (High Molecular Risk, "HMR"; international prognostic scoring system ("IPSS"); Dynamic IPSS ("DIPSS"). Outcomes assessed using overall survival ("OS") and AML transformation ("AML").

Diagnosis	Cohort	Model	Outcome	Time (yrs)	Brier	Absolute Prediction Error	Stderr	Uno C	Uncertainty
CP	Training-Xval	Full model	OS	5	0.05	0.11	0.02	0.86	0.18
CP	Training-Xval	Full model	OS	10	0.09	0.2	0.02	0.86	0.24
CP	Training-Xval	Full model	OS	15	0.09	0.23	0.02	0.86	0.11
CP	Training-Xval	Full model	AMLT	5	0.01	0.02	0.05	0.78	0.18
CP	Training-Xval	Full model	AMLT	10	0.01	0.03	0.06	0.78	0.24
CP	Training-Xval	Full model	AMLT	15	0.03	0.06	0.06	0.74	0.11
CP	Training-Xval	Full model	MFT	5	0.01	0.02	0.06	0.68	0.2
CP	Training-Xval	Full model	MFT	10	0.02	0.04	0.06	0.67	0.24
CP	Training-Xval	Full model	MFT	15	0.03	0.06	0.06	0.63	0.11
CP	Training-Xval	Full model	EFS	5	0.06	0.13	0.02	0.84	0.2
CP	Training-Xval	Full model	EFS	10	0.1	0.21	0.02	0.84	0.23
CP	Training-Xval	Full model	EFS	15	0.1	0.24	0.02	0.84	0.1
ET	Training-Xval	Full model	OS	5	0.06	0.12	0.02	0.86	0.17
ET	Training-Xval	Full model	OS	10	0.1	0.21	0.02	0.86	0.24
ET	Training-Xval	Full model	OS	15	0.09	0.24	0.02	0.86	0.1
ET	Training-Xval	IPSET	OS	-	-	-	0.02	0.72	-
ET	Training-Xval	Full model	AMLT	5	0.01	0.02	0.05	0.77	0.17
ET	Training-Xval	Full model	AMLT	10	0.02	0.03	0.05	0.76	0.24
ET	Training-Xval	Full model	AMLT	15	0.03	0.07	0.05	0.71	0.1
ET	Training-Xval	IPSET	AMLT	-	-	-	0.05	0.57	-
ET	Training-Xval	Full model	MFT	5	0.01	0.02	0.05	0.65	0.17
ET	Training-Xval	Full model	MFT	10	0.02	0.04	0.05	0.64	0.24
ET	Training-Xval	Full model	MFT	15	0.03	0.06	0.05	0.61	0.1

ET	Training-Xval	IPSET	MFT	-	-	-	0.05	0.6	-
ET	Training-Xval	Full model	EFS	5	0.07	0.14	0.02	0.83	0.18
ET	Training-Xval	Full model	EFS	10	0.1	0.23	0.02	0.84	0.24
ET	Training-Xval	Full model	EFS	15	0.1	0.25	0.02	0.84	0.09
ET	Training-Xval	IPSET	EFS	-	-	-	0.02	-	-
CP	External	Limited genomic data	OS	5	0.06	0.11	0.03	0.83	0.07
CP	External	Limited genomic data	OS	10	0.09	0.2	0.03	0.84	0.21
CP	External	Limited genomic data	OS	15	0.13	0.3	0.03	0.84	0.23
CP	External	Limited genomic data	AMLT	5	0.01	0.04	0.09	0.81	0.07
CP	External	Limited genomic data	AMLT	10	0.2	0.08	0.09	0.81	0.21
CP	External	Limited genomic data	AMLT	15	0.06	0.19	0.09	0.81	0.24
CP	External	Limited genomic data	MFT	5	0.03	0.06	0.04	0.58	0.09
CP	External	Limited genomic data	MFT	10	0.08	0.13	0.04	0.58	0.23
CP	External	Limited genomic data	MFT	15	0.13	0.23	0.04	0.56	0.21
CP	External	Limited genomic data	EFS	5	0.08	0.15	0.03	0.72	0.09
CP	External	Limited genomic data	EFS	10	0.14	0.27	0.03	0.72	0.23
CP	External	Limited genomic data	EFS	15	0.17	0.35	0.03	0.72	0.21

Appendix 6: Numbers needed to test using multistage model to detect one patient meeting required criteria (probability greater than threshold x%)

Group	Subgroup	Outcome	Timepoint (yrs)	Threshold (x%)	NNT
ET/PV	All	AML transformation	10	10	43
ET/PV	All	AML transformation	15	10	8
ET/PV	All	AML transformation	15	20	33
ET/PV	All	AML transformation	15	30	77
ET/PV	Under 60yrs	AML transformation	15	10	10
ET/PV	Under 60yrs	AML transformation	15	20	45
ET/PV	All	AML/MF transformation or Death	5	10	5
ET/PV	All	AML/MF transformation or Death	5	20	10
ET/PV	All	AML/MF transformation or Death	5	30	22
ET/PV	All	AML/MF transformation or Death	5	40	54
ET/PV	All	AML/MF transformation or Death	10	10	2
ET/PV	All	AML/MF transformation or Death	10	20	3
ET/PV	All	AML/MF transformation or Death	10	30	5
ET/PV	All	AML/MF transformation or Death	10	40	6
ET/PV	All	AML/MF transformation or Death	10	50	9
ET/PV	All	AML/MF transformation or Death	15	30	2
ET/PV	All	AML/MF transformation or Death	15	40	3
ET/PV	All	AML/MF transformation or Death	15	50	3
ET/PV	Under 60yrs	AML/MF transformation or Death	5	10	73
ET/PV	Under 60yrs	AML/MF transformation or Death	10	10	4
ET/PV	Under 60yrs	AML/MF transformation or Death	10	20	21
ET/PV	Under 60yrs	AML/MF transformation or Death	15	10	2
ET/PV	Under 60yrs	AML/MF transformation or Death	15	20	3
ET/PV	Under 60yrs	AML/MF transformation or Death	15	30	4
ET/PV	Under 60yrs	AML/MF transformation or Death	15	40	6
ET/PV	Under 60yrs	AML/MF transformation or Death	15	50	19
ET/PV	All	MF transformation	10	10	49
ET/PV	All	MF transformation	15	10	10
ET/PV	All	MF transformation	15	20	60
ET/PV	Under 60yrs	MF transformation	10	10	86
ET/PV	Under 60yrs	MF transformation	15	10	12
ET/PV	Under 60yrs	MF transformation	15	20	63
MF	All	AML transformation	5	10	4
MF	All	AML transformation	5	20	9
MF	All	AML transformation	5	30	17
MF	All	AML transformation	5	40	28
MF	All	AML transformation	5	50	56
MF	All	AML transformation	10	10	3
MF	All	AML transformation	10	20	6
MF	All	AML transformation	10	30	10
MF	All	AML transformation	10	40	17
MF	All	AML transformation	10	50	35
MF	Under 60yrs	AML transformation	5	10	13
MF	Under 60yrs	AML transformation	5	20	22
MF	Under 60yrs	AML transformation	5	30	55
MF	Under 60yrs	AML transformation	10	10	7
MF	Under 60yrs	AML transformation	10	20	13
MF	Under 60yrs	AML transformation	10	30	22
MF	Under 60yrs	AML transformation	10	40	37
MF	All	AML transformation or death	5	20	2
MF	All	AML transformation or death	5	30	3
MF	All	AML transformation or death	5	40	3
MF	All	AML transformation or death	5	50	4
MF	All	AML transformation or death	10	50	2
MF	Under 60yrs	AML transformation or death	5	10	2
MF	Under 60yrs	AML transformation or death	5	20	6
MF	Under 60yrs	AML transformation or death	5	30	10
MF	Under 60yrs	AML transformation or death	5	40	22
MF	Under 60yrs	AML transformation or death	5	50	55
MF	Under 60yrs	AML transformation or death	10	20	2
MF	Under 60yrs	AML transformation or death	10	30	3
MF	Under 60yrs	AML transformation or death	10	40	6
MF	Under 60yrs	AML transformation or death	10	50	8

ORIGINAL ARTICLE

Classification and Personalized Prognosis in Myeloproliferative Neoplasms

J. Grinfeld, J. Nangalia, E.J. Baxter, D.C. Wedge, N. Angelopoulos, R. Cantrill, A.L. Godfrey, E. Papaemmanuil, G. Gundem, C. MacLean, J. Cook, L. O'Neil, S. O'Meara, J.W. Teague, A.P. Butler, C.E. Massie, N. Williams, F.L. Nice, C.L. Andersen, H.C. Hasselbalch, P. Guglielmelli, M.F. McMullin, A.M. Vannucchi, C.N. Harrison, M. Gerstung, A.R. Green, and P.J. Campbell

ABSTRACT

BACKGROUND

Myeloproliferative neoplasms, such as polycythemia vera, essential thrombocythemia, and myelofibrosis, are chronic hematologic cancers with varied progression rates. The genomic characterization of patients with myeloproliferative neoplasms offers the potential for personalized diagnosis, risk stratification, and treatment.

METHODS

We sequenced coding exons from 69 myeloid cancer genes in patients with myeloproliferative neoplasms, comprehensively annotating driver mutations and copy-number changes. We developed a genomic classification for myeloproliferative neoplasms and multistage prognostic models for predicting outcomes in individual patients. Classification and prognostic models were validated in an external cohort.

RESULTS

A total of 2035 patients were included in the analysis. A total of 33 genes had driver mutations in at least 5 patients, with mutations in *JAK2*, *CALR*, or *MPL* being the sole abnormality in 45% of the patients. The numbers of driver mutations increased with age and advanced disease. Driver mutations, germline polymorphisms, and demographic variables independently predicted whether patients received a diagnosis of essential thrombocythemia as compared with polycythemia vera or a diagnosis of chronic-phase disease as compared with myelofibrosis. We defined eight genomic subgroups that showed distinct clinical phenotypes, including blood counts, risk of leukemic transformation, and event-free survival. Integrating 63 clinical and genomic variables, we created prognostic models capable of generating personally tailored predictions of clinical outcomes in patients with chronic-phase myeloproliferative neoplasms and myelofibrosis. The predicted and observed outcomes correlated well in internal cross-validation of a training cohort and in an independent external cohort. Even within individual categories of existing prognostic schemas, our models substantially improved predictive accuracy.

CONCLUSIONS

Comprehensive genomic characterization identified distinct genetic subgroups and provided a classification of myeloproliferative neoplasms on the basis of causal biologic mechanisms. Integration of genomic data with clinical variables enabled the personalized predictions of patients' outcomes and may support the treatment of patients with myeloproliferative neoplasms. (Funded by the Wellcome Trust and others.)

The authors' full names, academic degrees, and affiliations are listed in the Appendix. Address reprint requests to Dr. Green at the Cambridge Institute for Medical Research, Hills Rd., Cambridge CB2 0XY, United Kingdom, or at arg1000@cam.ac.uk; or to Dr. Campbell at the Wellcome Trust Sanger, Hinxton, Cambridgeshire CB10 1SA, United Kingdom, or at pc8@sanger.ac.uk.

Drs. Grinfeld and Nangalia and Drs. Green and Campbell contributed equally to this article.

N Engl J Med 2018;379:1416-30.
DOI: 10.1056/NEJMoa1716614

Myeloproliferative neoplasms are clonal hematopoietic disorders comprising polycythemia vera, which is characterized by red-cell overproduction; essential thrombocythemia, which involves elevated platelet counts; and myelofibrosis, which is defined by bone marrow fibrosis.[1] Polycythemia vera and essential thrombocythemia are chronic-phase myeloproliferative neoplasms, whereas myelofibrosis represents advanced disease that is diagnosed either initially or after the diagnosis of essential thrombocythemia or polycythemia vera. Current classification schemes distinguish among the subtypes of myeloproliferative neoplasms according to clinical and laboratory features,[2-5] but uncertainty clouds where and how to draw dividing lines among them.[6,7]

Biologically, the development of myeloproliferative neoplasms is driven by mutations in *JAK2*, *CALR*, or *MPL*. Many patients have additional drivers that span a wide range of cancer genes, with patient-to-patient variation in the genetic and clonal landscape.[8,9] Driver mutations correlate with phenotype and prognosis,[10-12] and mutation order can also influence phenotype.[13,14] This complex genetic landscape probably contributes to heterogeneity in diagnostic features and outcomes in patients with myeloproliferative neoplasms.

In blood cancers, a progressive shift is under way, from clinical and morphologic classification schemes to those that are based on genomics.[15] Driver mutations are increasingly important in predicting clinical outcomes, but large, well-characterized cohorts are necessary for accurate prognostic models.[16] Studies have suggested that this promise extends to myeloproliferative neoplasms,[10,17] but larger cohorts and comprehensive gene sequencing are needed in order to provide definitive answers.

METHODS

STUDY SAMPLES

We analyzed samples that were obtained from patients after they provided written informed consent and after ethics approval from relevant authorities was obtained. Details regarding the cohort, disease classification, and diagnostic review are provided in the Supplementary Appendix, available with the full text of this article at NEJM.org. Tumor DNA was derived from blood granulocytes, bone marrow mononuclear cells, or whole blood. The majority of patients did not have matched germline samples sequenced. We use the term "myelofibrosis" to encompass both primary myelofibrosis and myelofibrosis that evolved from essential thrombocythemia or polycythemia vera.

No commercial support was involved in this study. See the Supplementary Appendix for details regarding patient cohorts and samples.

SEQUENCING AND ANALYSES

We designed custom RNA baits to capture the full coding sequence of 69 genes, single-nucleotide polymorphisms for copy number profiling, and germline loci that have been associated with myeloproliferative neoplasms (Tables S1 and S2 in the Supplementary Appendix).[18-20] Additional patients underwent whole-exome sequencing, as reported previously.[8]

CLINICAL VARIABLES

Laboratory and clinical data from diagnosis were incorporated into prognostic models. The median duration between diagnosis and sample acquisition was 49 days. The median follow-up was 93.8 months (range, <1 to 523) from diagnosis and 72.0 months (range, <1 to 360) from DNA sampling.

STATISTICAL ANALYSIS

We estimated the timing of mutation acquisition using Bradley–Terry models of pairwise comparisons of clonal fractions.[13] We used a Bayesian network analysis and Dirichlet processes to identify genetic associations and subgroups. Random-effects Cox proportional-hazards multistate modeling was used for outcome predictions (see the Supplementary Appendix).

RESULTS

SPECTRUM OF GENOMIC CHANGES IN MYELOPROLIFERATIVE NEOPLASMS

Targeted sequencing for the full coding sequence of 69 genes and genomewide copy-number information was undertaken in 1887 patients, and 148 patients underwent whole-exome sequencing, as reported previously.[8] The cohort of 2035 patients included 1321 patients with essential thrombocythemia, 356 with polycythemia vera, 309 with myelofibrosis, and 49 with other diagnoses of myeloproliferative neoplasms (Table S3 in the Supplementary Appendix). A total of 33 genes had driver mutations in at least 5 patients (Fig. 1A,

and Tables S4 and S5 in the Supplementary Appendix). Mutations in *JAK2*, *MPL*, and *CALR* accounted for 1831 driver mutations and were the sole abnormality in 45% of the patients. A total of 1075 driver mutations were identified across other genes. Loss of heterozygosity was frequent for *JAK2* V617F, especially in patients with polycythemia vera, but was infrequent for *CALR* and *MPL* (Fig. S1 in the Supplementary Appendix).

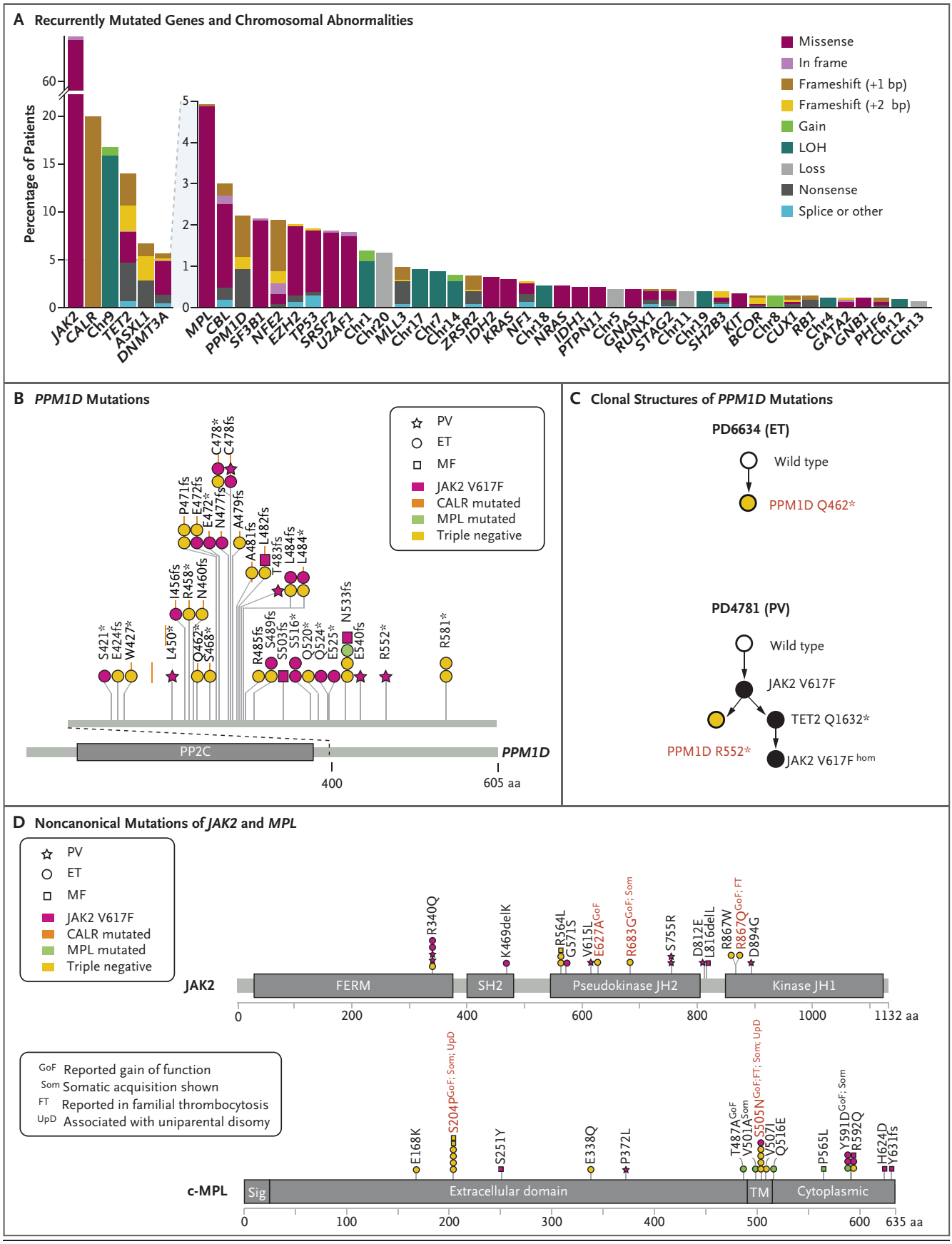
We identified 45 truncating mutations in the terminal exon of *PPM1D* in 38 patients within the cohort (1.9%) (Fig. 1B); thus, *PPM1D* was the eighth most commonly mutated gene in myeloproliferative neoplasms. These mutations have also been detected in solid tumors, blood samples obtained from healthy persons, and patients with breast or ovarian tumors, often after chemotherapy.[21,22] In our cohort, 10 patients had *PPM1D* mutations that were detectable only in a later sample obtained during treatment with hydroxyurea. However, *PPM1D* mutations were also detected at or within 1 month after diagnosis in 20 patients. Analysis of single-cell–derived hematopoietic colonies identified mutated *PPM1D* in a patient with triple-negative essential thrombocythemia (i.e., nonmutated *JAK2*, *CALR*, or *MPL*) but also identified mutated *PPM1D* that was subclonal to *JAK2* V617F in a patient with polycythemia vera (Fig. 1C). These data confirm that *PPM1D* mutations can occur within the myeloproliferative neoplasm clone and be present at diagnosis; thus, their presence does not always indicate age-related clonal hematopoiesis or therapy-related disease evolution.

Mutations in *MLL3* were detected in 20 patients (1.0%) and were predominantly nonsense or frameshift, as has been reported in patients with acute myeloid leukemia (Fig. 1A, and Table S4 in the Supplementary Appendix).[23] Among these 20 patients, 7 patients had triple-negative myeloproliferative neoplasms, which suggests that *MLL3* could be an important tumor-suppressor gene in these patients.

Whether mutations in *JAK2* and *MPL* outside the known hot spots could be relevant to patients with myeloproliferative neoplasms has been unclear.[24,25] We identified noncanonical variants in *JAK2* and *MPL* in 16 patients with triple-negative essential thrombocythemia and in 1 patient with triple-negative myelofibrosis (Fig. 1D). Of these, three groups of variants were likely to be relevant to disease pathogenesis: *JAK2* R683G and *JAK2* E627A in 2 patients with essential thrombocythemia (reported in acute lymphoblastic leukemia in which they activate JAK2[26-28]); *JAK2* R867 in 2 patients with essential thrombocythemia (associated with familial thrombocythemia[29]); and *MPL* S505N and *MPL* S204P in 4 and 5 patients, respectively, with essential thrombocythemia.[24] *MPL* S204P co-occurred with loss of heterozygosity (LOH) at chromosome 1p, which suggests

Figure 1 (facing page). Genomic Landscape of Myeloproliferative Neoplasms.

Panel A shows the frequency of recurrently mutated genes and chromosomal abnormalities in the cohort of 2035 patients. Mutations were stratified according to type (missense, nonsense, affecting a splice site, or other [e.g., stop, gain, or loss]). Insertions and deletions (del) were categorized according to whether they resulted in a shift in the codon reading frame (by either 1 or 2 base pairs [bp]) or were in frame. Chromosomes are indicated by Chr plus a numeral (e.g., Chr9 denotes chromosome 9). Chromosomal gains include whole-chromosome gains (trisomy) and subchromosomal amplifications. Chromosomal losses include whole-chromosome deletions (monosomy) and subchromosomal deletions. Loss of heterozygosity (LOH) was predominantly copy-number neutral, but in some cases, chromosome losses could not be ruled out. Panel B shows the site within the gene and protein consequence of *PPM1D* mutations. Colored shapes represent the characteristics of the patient who had the specific mutation (shapes indicate the subtype of myeloproliferative neoplasm, and colors the phenotypic driver). A triple-negative finding indicates nonmutated *JAK2*, *CALR*, and *MPL*. The term aa denotes amino acid, ET essential thrombocythemia, fs frameshift, MF myelofibrosis, PP2C protein phosphatase 2C domain, and PV polycythemia vera. Panel C shows clonal structures of two patients with *PPM1D* mutations determined by genotyping of hematopoietic colonies derived from peripheral-blood mononuclear cells. Each circle represents a group of hematopoietic colonies that share the same genotype: wild type (white), other driver mutations (black), and *PPM1D* mutated (yellow). Wild-type colonies are represented at the top of each diagram, with subsequent mutant subclones shown below. Somatic mutations acquired in each subclone are indicated beside respective nodes and represent those that were acquired in addition to mutations present in earlier subclones. The term hom denotes homozygous. Panel D shows the site within the gene and protein consequence of noncanonical mutations of *JAK2* and *MPL*. The V617F and exon 12 mutations in *JAK2* and W515 mutations in *MPL* are not shown. Mutations highlighted in red are likely to be relevant to disease pathogenesis, with previous studies having shown somatic acquisition, familial inheritance, or functional consequences for the specific variants (see box of abbreviations). FERM denotes the 4.1–ezrin–radixin–moesin domain, SH2 Src homology 2, Sig signal, and TM transmembrane.



a clonal advantage to acquired homozygosity for this variant.

FACTORS INFLUENCING CLASSIFICATION INTO DISEASE SUBTYPES

Currently, patients with myeloproliferative neoplasms are classified as having essential thrombocythemia, polycythemia vera, or myelofibrosis on the basis of clinical and laboratory criteria,[2-5] but the biologic factors underlying these distinctions are incompletely understood. The number of driver mutations per patient was higher in those with myelofibrosis than in those with polycythemia vera or essential thrombocythemia (Fig. 2A), as previously reported,[8] and increased according to the age of the patient (Fig. 2B).

The distinction between *JAK2* V617F–mutated essential thrombocythemia and polycythemia vera rests on whether the red-cell mass or hematocrit is elevated. We found that acquired driver mutations correlated with hematologic variables (Fig. S2 in the Supplementary Appendix) and were the strongest determinants of a patient with *JAK2* V617F–mutated chronic-phase disease receiving a diagnosis of essential thrombocythemia as compared with polycythemia vera, although germline genetic background and demographic factors also contributed (Fig. 2C, and Fig. S2 in the Supplementary Appendix). LOH at chromosome 9p (9pLOH), causing *JAK2* V617F homozygosity, or a high *JAK2* V617F allele burden correlated with polycythemia vera, as did mutated *NFE2*, a transcription factor critical to erythroid differentiation.

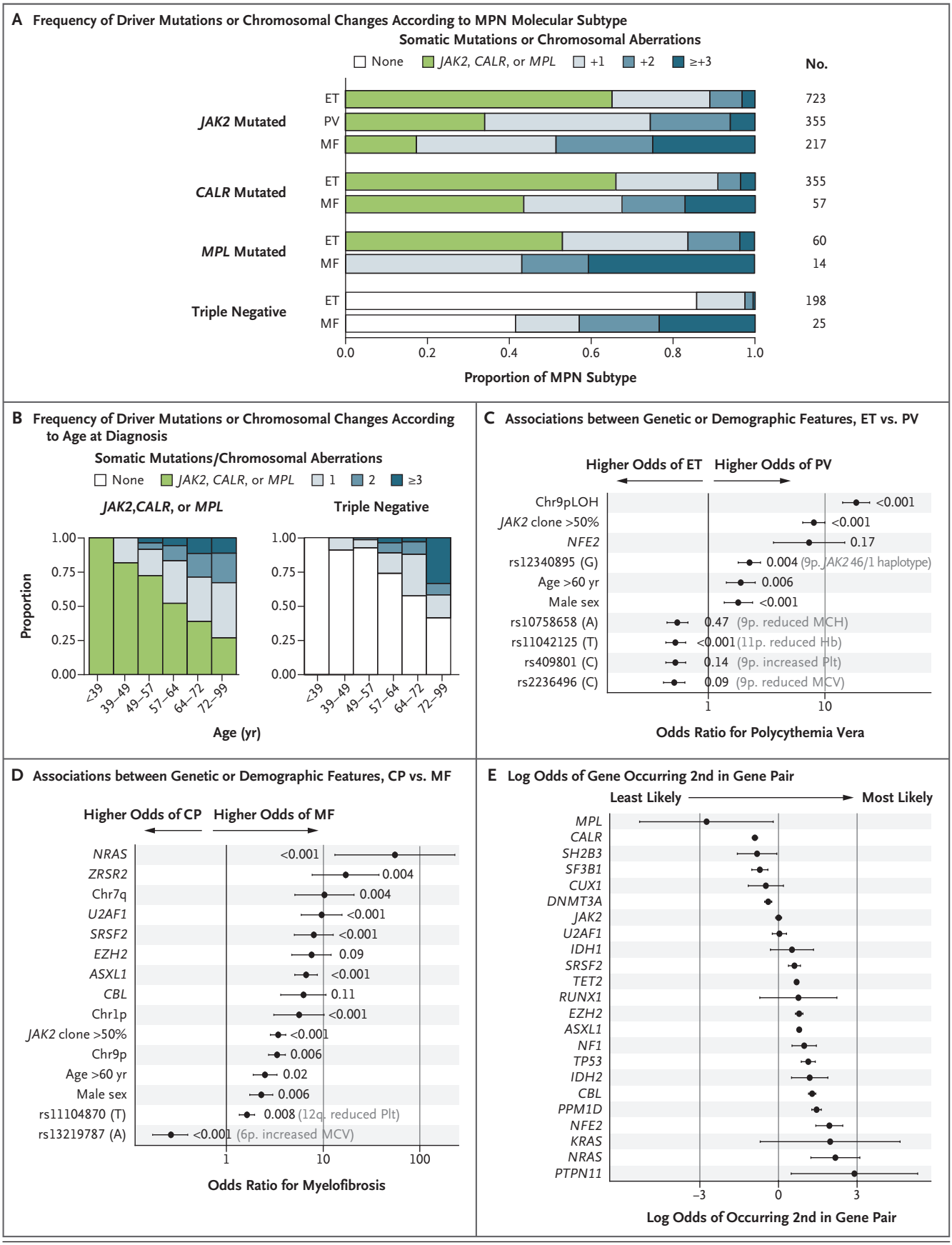
Germline polymorphisms that have been associated with red-cell variables in the general population were distributed unevenly, with alleles associated with lower hemoglobin level and higher platelet counts being enriched in patients with essential thrombocythemia (Fig. 2C). Furthermore, the *JAK2* 46/1 haplotype, which is known to increase the predisposition to myeloproliferative neoplasms,[18] correlated with polycythemia vera (odds ratio, 2.3; 95% confidence interval [CI], 1.7 to 3.3; P=0.004), possibly through increasing odds of *JAK2* V617F homozygosity by 9pLOH (odds ratio, 2.7; 95% CI, 2.0 to 3.9; P<0.001). Older age and male sex also increased the odds of polycythemia vera. These data show that the location of any chronic-phase disease on the hemoglobin and red-cell mass continuum is influenced by many factors and that any arbitrary threshold to label patients' disease as being one subtype or the other will not distinguish among patients with different underlying biologic factors.

Figure 2 (facing page). Factors Affecting Disease Classification at Presentation and Timing of Somatic Mutations.

Histograms show the frequency of driver mutations or chromosomal changes (gains, losses, or LOH) that were identified in different molecular subtypes of myeloproliferative neoplasm (MPN) (excluding 24 patients with >1 detectable phenotypic driver mutation) (Panel A) and according to the age of the patient at diagnosis (Panel B). Forest plots showing the associations between genetic or demographic features and presentation with essential thrombocythemia (ET) as compared with polycythemia vera (PV) in patients with *JAK2* V617F mutations (Panel C) and the presentation in chronic-phase (CP) disease as compared with MF across patients with *JAK2*, *CALR*, or *MPL* mutations (Panel D). Significant associations from univariate analyses after correction for multiple hypothesis testing are shown. P values were derived from logistic-regression modeling, identifying independent associations. Causes of reductions in the hemoglobin (Hb) level, mean corpuscular hemoglobin (MCH) level, mean corpuscular volume (MCV), and platelet (Plt) count are indicated. Of 671 patients who had more than one somatic mutation, the order of mutation acquisition of at least one pair of mutations was determined in 271 patients (40%) (Panel E). These ordered pairings were used to determine the relative probabilities of a gene occurring first or second for a given pairing with the use of Bradley–Terry modeling, which provided an estimate of the overall timing of mutation acquisition. The horizontal axis shows the log odds of a gene occurring second in a gene pair. For example, as compared with *JAK2*, *PPM1D* mutations have a log odds of 1.45 and therefore are $e^{1.45}$, or 4.3, times more likely to occur second in the pair. Any pair of genes can be assessed in this manner by calculating the exponential of the difference in log odds for gene A and gene B. The error bars indicate 95% confidence intervals.

Mutations in spliceosome components, epigenetic regulators, and the RAS pathway were strongly associated with accelerated phase (myelofibrosis), as compared with chronic-phase disease (essential thrombocythemia or polycythemia vera), as were male sex, older age, and germline loci associated with platelet count and red-cell variables (Fig. 2D).

The order in which mutations are acquired in myeloproliferative neoplasms has previously been shown to influence disease phenotype.[13,14] *CALR* and *MPL* mutations occurred more commonly early in disease, whereas mutations in *NRAS*, *TP53*, *PPM1D*, and *NFE2* were acquired significantly later in disease (Fig. 2E, and Fig. S3 in the Supplementary Appendix). Some of the earlier-occurring mutations in genes such as *SF3B1* and



DNMT3A are also associated with age-related clonal hematopoiesis,[30,31] which suggests that some myeloproliferative neoplasms could arise from an antecedent asymptomatic clone. In patients with multiple mutations, *JAK2* V617F was more commonly a secondary event in patients with essential thrombocythemia and an earlier event in those with polycythemia vera or myelofibrosis (Figs. S4 and S5 in the Supplementary Appendix), a finding that confirms and generalizes observations that had previously been shown for *JAK2* relative to *TET2* or *DNMT3A*.[13,14]

GENOMIC SUBGROUPS IN MYELOPROLIFERATIVE NEOPLASMS

Hematologic cancers may be subclassified according to driver mutations that distinguish subgroups of patients,[32,33,34] with the use of patterns of mutually exclusive or co-mutated genes. In our cohort, driver mutations showed complex patterns of assortment (Fig. S6 in the Supplementary Appendix). We used Bayesian modeling to identify genomic subgroups of myeloproliferative neoplasms with maximum within-group similarity and maximum between-group discrimination.

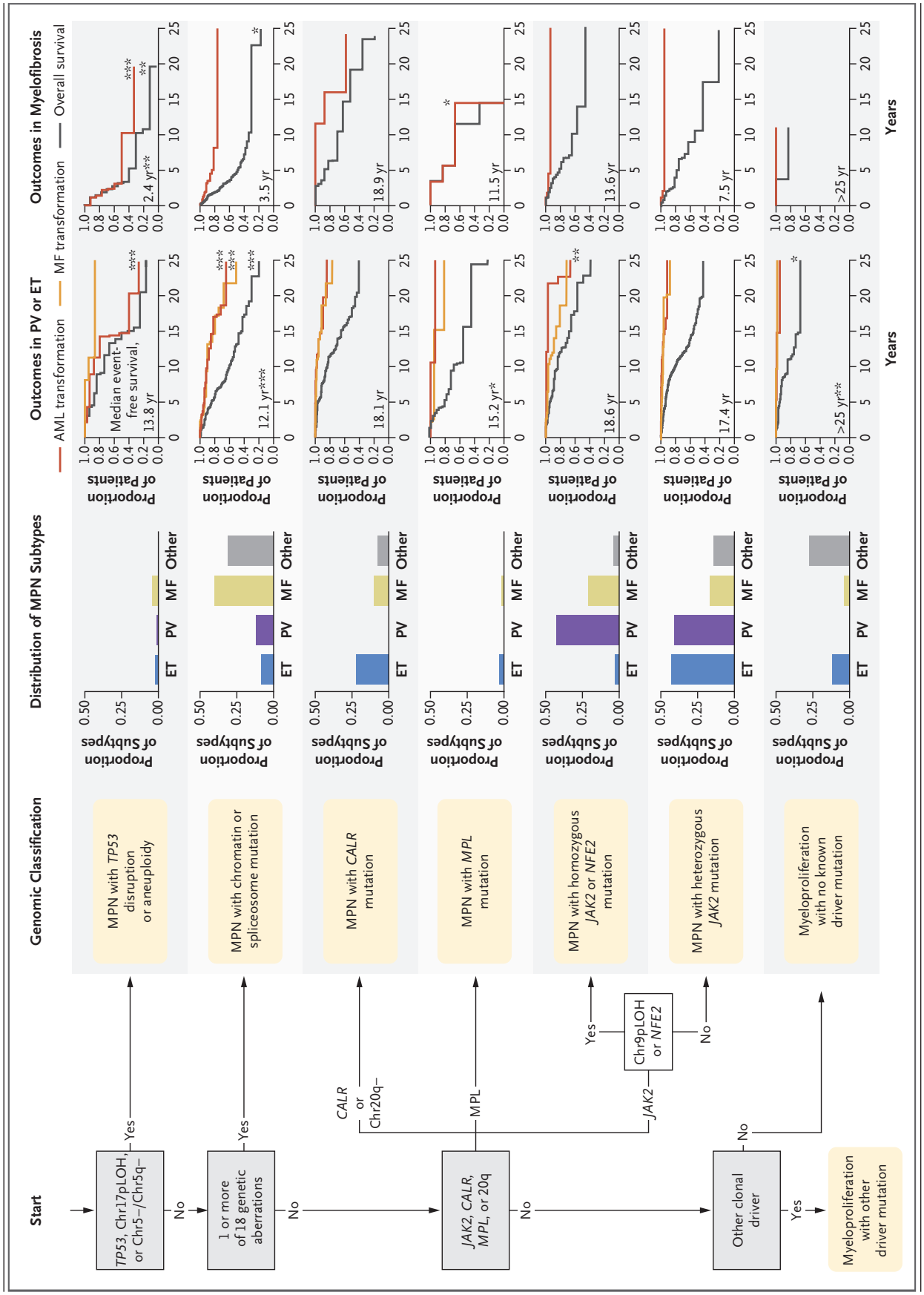
We identified eight genomic subgroups in myeloproliferative neoplasms, defined according to simple rules (Fig. 3, and Fig. S7 in the Supplementary Appendix). *TP53* mutations, often co-occurring with aberrations at chromosome 17p, and deletions at chromosome 5q identified the first subgroup. *TP53* mutations often occur later in disease (Fig. 2E) but dominate the genomic and clinical features of these patients regardless of the initial driver of the myeloproliferative neoplasm. As in patients with other blood cancers with *TP53* mutations,[32,35] these patients have a dismal prognosis with a high risk of transformation to acute myeloid leukemia (hazard ratio vs. the *JAK2*-heterozygous subgroup, 15.5; 95% CI, 7.5 to 31.4; P<0.001) and early death (hazard ratio, 2.4; 95% CI, 1.6 to 3.6; P<0.001).

The second subgroup was defined by the presence of one or more mutations in 16 myeloid cancer genes, especially chromatin and spliceosome regulators, LOH at chromosome 4q, and aberrations in chromosomes 7 and 7q. This subgroup was enriched for patients with myelofibrosis (odds ratio, 6.5; 95% CI, 4.9 to 8.7; P<0.001) and myelodysplastic–myeloproliferative neoplasms (including all seven patients with chronic myelomonocytic leukemia or atypical chronic myeloid leukemia) but also included 8.4% of patients with essential thrombocythemia and 11.5% of those with polycythemia vera. Patients were at increased risk for transformation to myelofibrosis (hazard ratio vs. the *JAK2*-heterozygous subgroup, 5.4; 95% CI, 2.7 to 11.0; P<0.001) and shorter event-free survival, regardless of myeloproliferative neoplasm subtype or phenotypic driver mutation (hazard ratio for disease progression or death, 2.6; 95% CI, 2.1 to 3.2; P<0.001).

Patients who were not identified in the above two subgroups were classified according to their dominant myeloproliferative neoplasm phenotypic driver mutation. Patients with *CALR* mutations, which co-occurred with LOH at chromosome 19p and with deletion at chromosome 20q, or those with *MPL* mutations all presented with essential thrombocythemia or myelofibrosis. Patients with *MPL*-mutated myelofibrosis had an elevated rate of acute myeloid leukemia transformation (hazard ratio vs. the *JAK2*-heterozygous subgroup, 8.6; 95% CI, 1.4 to 49.1; P=0.02), but otherwise the two subgroups had a clinical course that was similar to that in the *JAK2* subgroups. Patients with *JAK2* V617F heterozygosity constituted most of the patients with *JAK2*-

Figure 3 (facing page). Genomic Subgroups in MPN and Phenotypic Characteristics.

According to a Bayesian clustering algorithm (Dirichlet process), patients could be classified into six distinct subgroups on the basis of the presence or absence of mutations and chromosomal abnormalities. The remaining patients either had no detectable genomic changes or had clonal markers that were not defining for one of the six groups. The flowchart shows the logic that allows patients to be classified into the total of eight groups. Proportions of patients with essential thrombocythemia (ET), polycythemia vera (PV), myelofibrosis (MF, either primary or after chronic-phase disease), or other MPN diagnoses are shown, as are rates of overall survival and myelofibrotic or leukemic transformation among patients in the individual subgroups. The 18 genetic aberrations involved *EZH2*, *IDH1*, *IDH2*, *ASXL1*, *PHF6*, *CUX1*, *ZRSR2*, *SRSF2*, *U2AF1*, *KRAS*, *NRAS*, *GNAS*, *CBL*, Chr7/7qLOH, Chr4qLOH, *RUNX1*, *STAG2*, and *BCOR*. Patients who had more than one mutation across *JAK2*, *CALR*, and *MPL* and deletion at chromosome 20q could belong to more than one classification. In patients who had myeloproliferation with other driver mutations, other diagnoses should be considered, depending on the nature of the genetic aberration. Chromosome 9pLOH was judged to be present if detectable at a 10% clonal fraction. The number of asterisks indicates the P value (*P<0.05, **P<0.01, and ***P<0.001) for the comparison with patients with MPN with heterozygous *JAK2* mutation.



mutated essential thrombocythemia but also some of the patients with polycythemia vera or myelofibrosis; these patients had generally favorable outcomes. The subgroup of patients with *JAK2* homozygosity was enriched for patients with *NFE2* mutations and for patients with polycythemia vera. Myelofibrosis transformations occurred more frequently in this subgroup (hazard ratio vs. the *JAK2*-heterozygous subgroup, 3.0; 95% CI, 1.3 to 6.6; P=0.007).

A seventh subgroup (36 patients [1.8%]) had identifiable driver mutations but none of the class-defining drivers identified above. This included patients with mutations in genes such as *TET2* and *DNMT3A* that are not disease-specific or with mutations in genes that have been associated with other myeloid cancers (such as *KIT* in systemic mastocytosis). The eighth subgroup (192 patients [9.4%]) had no detectable driver mutations and may have included patients with either reactive thrombocythemia or myeloproliferative neoplasms with unidentified drivers. Patients were typically young and female and had received a diagnosis of essential thrombocythemia. This subgroup had particularly benign outcomes; only 1 patient (0.5%) had myelofibrosis transformation and 2 (1%) had acute myeloid leukemia transformation during a median follow-up of 8.0 years (hazard ratio for disease progression or death vs. the *JAK2*-heterozygous subgroup, 0.56; 95% CI, 0.38 to 0.78; P=0.005).

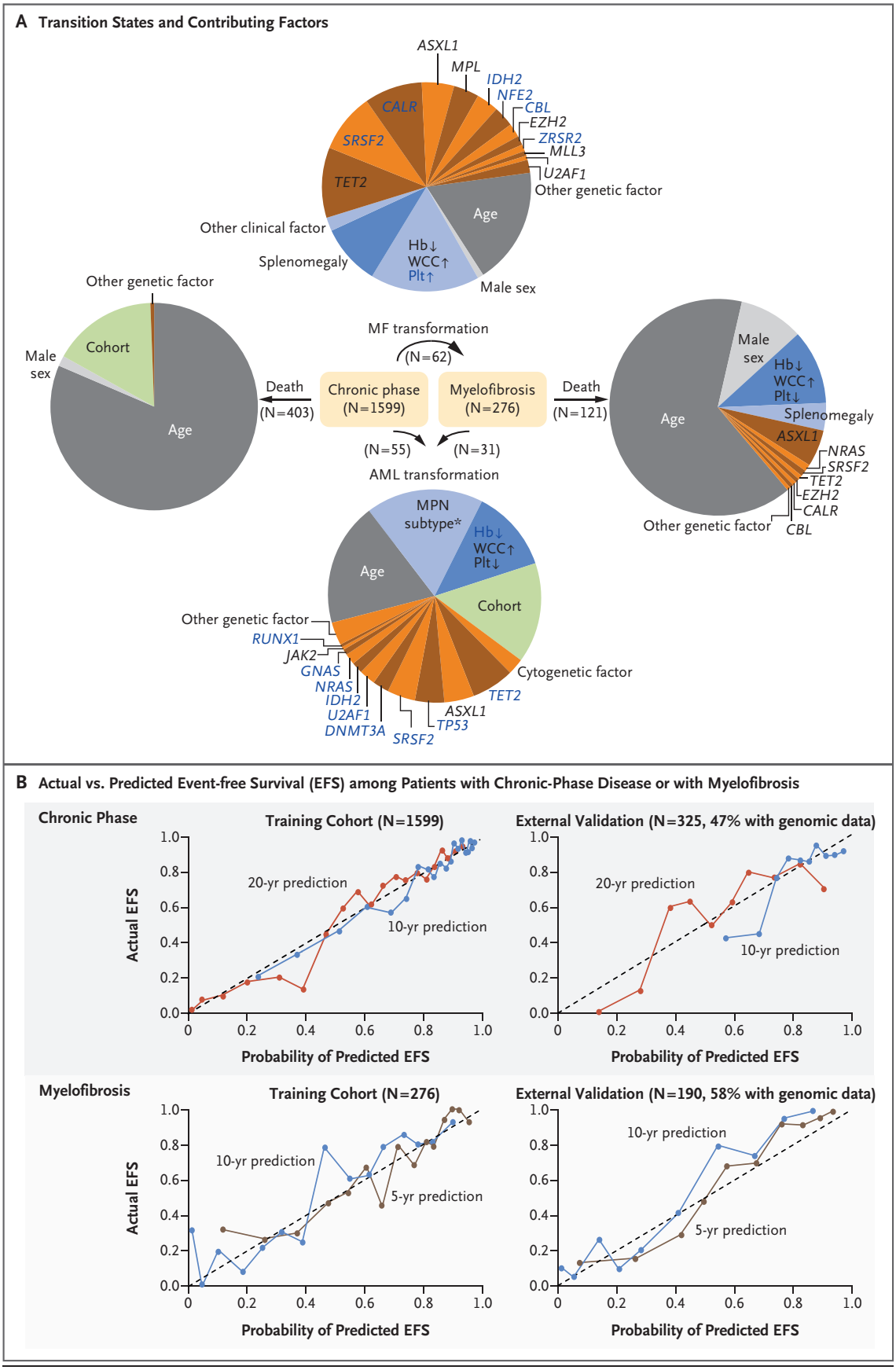
We applied our proposed classification scheme to an external cohort of 270 patients with myeloproliferative neoplasms (137 patients with essential thrombocythemia, 14 with polycythemia vera, and 119 with myelofibrosis) that had sufficient genomic characterization so that our flowchart could be applied. The subgroup proportions were similar in the two cohorts (Fig. S7 in the Supplementary Appendix).

Figure 4 (facing page). Modeling Outcome in Patients.
Panel A shows the transition states during a patient's disease and the factors contributing to the risk of each transition. Patients may have presented with either chronic-phase disease (polycythemia vera, essential thrombocythemia, or unclassifiable MPN) or myelofibrosis (MF), as represented by the two central, rounded rectangles. The patient may have subsequently remained alive in these disease states or, alternatively, could have transitioned to one of four states: death in chronic-phase disease, death in MF, MF transformation of chronic-phase disease, and acute myeloid leukemia (AML) transformation of either chronic-phase disease or MF. Individual models were created for each of these four disease-state transitions and combined into a single multistate model allowing for the prediction of probability of being in each disease state occurring at any time point in the future (up to 25 years after diagnosis), as calculated on an individual patient basis. Pie charts show the variables that contributed most to the predicted risk for each of the four transitions. These show the effect on disease transitions of both rare variables with a strong effect and common variables with a milder effect. Variables with a hazard ratio of more than 2.0 are shown in blue type. The numbers of patients with chronic-phase disease or MF are shown alongside the numbers of patients who transitioned to other states. Patients may have transitioned more than once during their clinical course (e.g., from chronic-phase disease to MF and then to AML). The risk of AML transformation was highest among patients with MF. WCC denotes white-cell count; the arrows by the clinical variables indicate whether the value increased (up arrow) or decreased (down arrow). Panel B shows the model predictions, as compared with the actual event-free survival (EFS), among patients. Comparisons of the actual EFS with the predicted EFS derived from multistate random-effects Cox proportional-hazards modeling for patients with chronic-phase disease and MF, for both the training cross-validation cohort and the external validation cohort, are shown. Each cohort was split into equally sized subgroups of patients, and each subgroup is represented by a data point plotted according to the observed and predicted EFS. Overall, the models show good correlation between predicted and actual outcomes for both the training and external validation cohorts at several time points (brown indicates the EFS at 5 years, blue at 10 years, and red at 20 years). The dashed line indicates points at which predicted outcomes perfectly match observed outcomes.

FACTORS INFLUENCING DISEASE PROGRESSION

A key determinant of the treatment of patients with myeloproliferative neoplasms is the predicted prognosis. For example, patients who are expected to have a benign future clinical course would probably benefit from treatments that are aimed at minimizing thrombotic risk, and those who are expected to have progression to leukemia or myelofibrotic bone marrow failure could be candidates for intensive therapy or clinical trials of new agents. We developed multivariate statistical models, incorporating 63 clinical and genomic variables, that estimated a patient's probability of transition between stages of disease — namely, chronic-phase disease (essential thrombocythemia or polycythemia vera), advanced-phase disease (myelofibrosis), acute myeloid leukemia, and death.

We determined the fraction of explained variation for each outcome that was attributable to different prognostic factors (Fig. 4A). Death in



the chronic phase was influenced predominantly by age, with genomic features having little predictive power — a finding that suggests that once cytoreduction has achieved adequate control of blood counts, causes of death are dominated by those that would also occur in the general population.[36] These would, therefore, not be well predicted by the specific genomic features of the myeloproliferative neoplasm.

By contrast, genomic features played a substantial role in predicting progression from chronic-phase disease to myelofibrosis and to acute leukemia transformation (Fig. 4A). *CALR* mutations were independently associated with an increased risk of myelofibrotic transformation, as previously reported.[37] Mutations in epigenetic regulators, splicing factors, and RAS signaling were all associated with myelofibrotic and leukemic transformation — some of these associations have been identified previously.[10-12] Whether mutations were clonal or subclonal had little effect on prognosis (see the Supplementary Appendix). Clinical features of the disease, such as anemia, splenomegaly, or thrombocytosis, still retained independent predictive power for transformation events, which suggests that these variables reflect important features of the disease state that are not captured in the genomic landscape. Outcomes in patients with myelofibrosis did not significantly differ on the basis of whether the myelofibrosis was primary or occurred after essential thrombocythemia or polycythemia vera.

Figure 5 (facing page). Personalized Predictions of Outcomes in Patients.

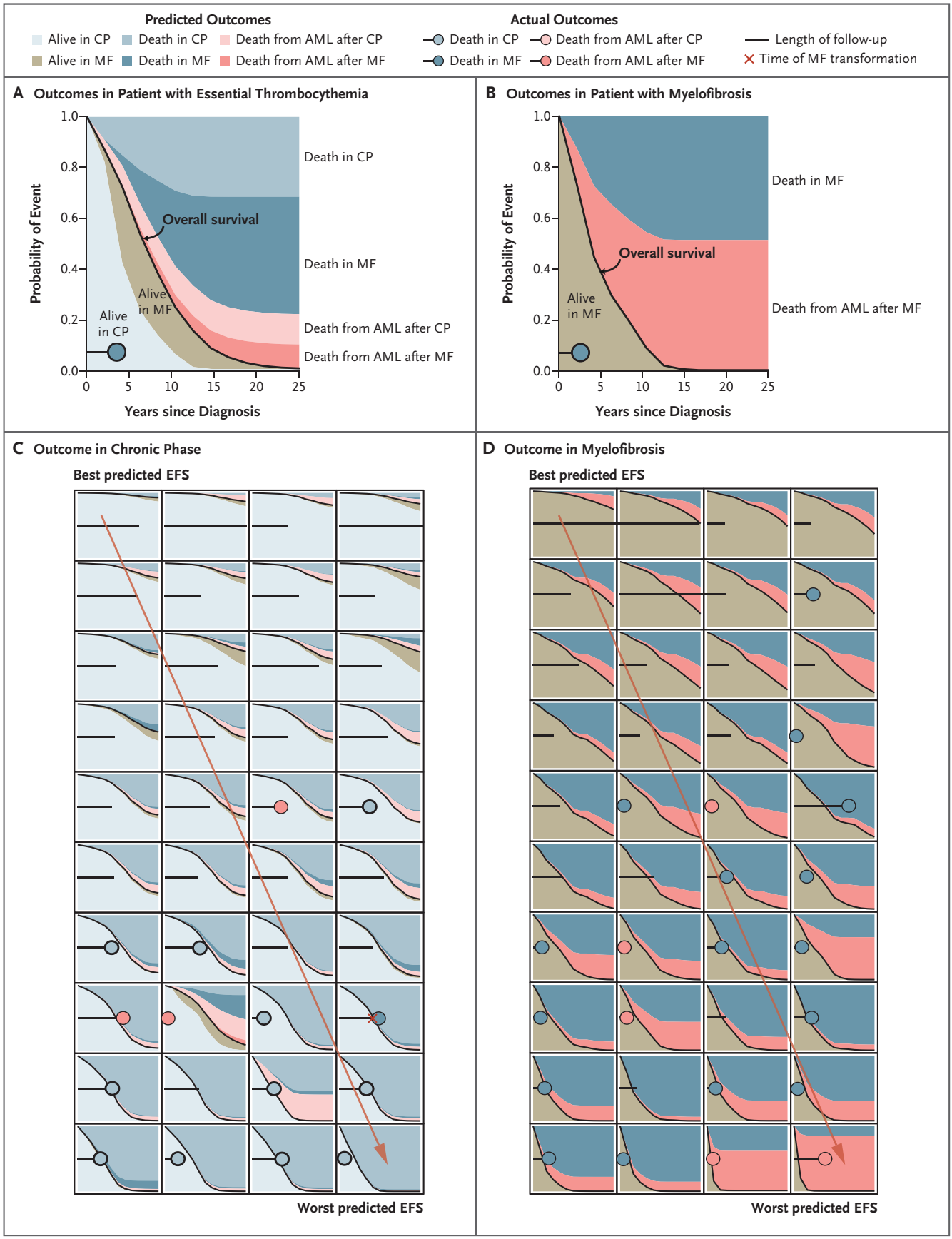
Panels A and B show example tiles that represent personalized predicted outcomes in individual patients. Panel A shows the predicted outcomes of a 79-year-old woman who presented with essential thrombocythemia (ET) with a hemoglobin level of 104 g per liter, a white-cell count of 8400 per cubic millimeter, and a platelet count of 2,300,000 per cubic millimeter, and mutated *CALR*, *SRSF2*, and *IDH2* along with LOH in chromosome 18q. For such a patient presenting with chronic-phase (CP) disease (PV or ET), the model incorporates all clinical, demographic, laboratory, and genomic variables to predict the overall probabilities over time of being alive in CP, dying in CP, being alive in myelofibrosis (MF) after CP, dying in MF after CP, transitioning to AML from CP, and transitioning to AML from MF after CP. The varying probabilities of each of these transitions can be judged from the vertical axis and their respective Kaplan–Meier curves over a 25-year period shown along the horizontal axis. The black curve shows the predicted Kaplan–Meier curve of overall survival. This patient transitioned to MF and died within 5 years; this outcome is shown along the bottom of the plot, where the length of the horizontal black line shows the duration of follow-up and the cause of death is indicated by the shading of the circle. For a patient who presented with MF, as shown in Panel B, the same model predicts the probabilities of being alive in MF, dying in MF, and transitioning to AML over a period of 25 years. Panel B shows the predicted and actual outcomes of a 57-year-old man with MF who had a hemoglobin level of 125 g per liter, a white-cell count of 27,000 per cubic millimeter, and a platelet count of 119,000 per cubic millimeter, and mutated *TET2*, *ASXL1*, *CBL*, and *BCOR* along with deletion in chromosomes 7q and 11q. This patient died in MF within 2 years. All patients with chronic-phase disease or MF who had either a disease event (death or disease progression) or had more than 10 years of follow-up (>5 years for patients with MF) were ranked according to their overall predicted EFS. The predicted and actual outcomes for 40 individual patients with CP (Panel C) and MF (Panel D) showing how patients in the cohort may be distinguished in terms of EFS and cause of death.

PERSONALLY TAILORED PROGNOSIS

Current prognostic models for myeloproliferative neoplasms, which are focused on myelofibrosis, use simple scoring systems and group patients into broad prognostic categories. Many factors influence clinical outcomes, with a wide range of effect sizes, which means that current schemes discard information that is relevant to prognosis. We explored whether our multivariate, multistate prognostic models could generate accurate predictions for individual patients.

The usefulness of personally tailored predictions can be assessed in two ways: do the predictions usefully distinguish among patients according to prognosis, and are the predictions more informative than conventional schemas? Regarding the first question, not only is our model able to generate a wide range of specific risk predictions (regarding long-term survival, death in chronic-phase disease, and myelofibrotic and leukemic transformation) but they correlate well with observed outcomes (Figs. 4B and 5, and Fig. S8 and Tables S6 and S7 in the Supplementary Appendix), both in cross-validation of an internal cohort and in an external validation cohort of 515 patients with myeloproliferative neoplasms (137 patients with essential thrombocythemia, 188 with polycythemia vera, and 190 with myelofibrosis).

Internal cross-validation showed concordances of 76 to 86% for overall survival, event-free



survival, and transformation to acute leukemia as well as good performance on absolute predictive accuracy (Fig. 4B, and Tables S6 and S7 in the Supplementary Appendix). Concordances were similar in the external cohort, despite the fact that patients in the external cohort received diagnoses at another center, were evaluated by different pathologists who used different diagnostic criteria, and underwent sequencing at a different facility with the use of a different gene panel from the training cohort (Fig. 4B). Thus, the model provides considerable discriminatory power that accurately generalizes to other real-world cohorts. Owing to the existence of different diagnostic criteria, the model does not rely heavily on the exact classification label of the patient's disease. Indeed, removing the distinction between polycythemia vera and essential thrombocythemia, but simply retaining the distinction between myelofibrosis and chronic-phase disease, did not reduce the predictive accuracy of the model (Fig. S9 in the Supplementary Appendix).

Our model showed superior performance to current major prognostic schemas in clinical use, such as the International Prognostic Scoring System (IPSS),[38] the Dynamic IPSS (DIPSS),[39] the high molecular risk category for myelofibrosis,[10] and the International Prognostic Score for Essential Thrombocythemia score[40] (Fig. S9 and Tables S6 and S7 in the Supplementary Appendix). Furthermore, we identified substantial heterogeneity in disease outcomes within individual prognostic categories of current prognostic schemas (shown for DIPSS in Fig. S10 in the Supplementary Appendix); this was especially prominent for intermediate-risk patients and allowed for more informative predictions in a group with otherwise uncertain outcomes. This means that not so many patients need be screened before some emerge as having an increased risk of poor outcomes; the numbers needed to test across different scenarios are shown in Table S8 in the Supplementary Appendix. The inclusion of mutations and chromosomal changes beyond *JAK2*, *CALR*, and *MPL* improved the predictive power of prognostic models (Tables S6 and S7 in the Supplementary Appendix).

We have implemented a free, user-friendly online calculator of individualized patient outcomes (https://cancer.sanger.ac.uk/mpn-multistage/) that enables the exploration of data from patients in our cohort, and the generation of new patient predictions according to available clinical, laboratory and genomic features. Further validation of our model with the use of additional cohorts of patients with myeloproliferative neoplasms will be important, given the bias toward including patients with essential thrombocythemia in this study.

DISCUSSION

A major challenge is how we use our understanding of the pathogenic complexity of myeloproliferative neoplasms to identify groups of patients with shared causative biologic factors of disease, such that existing and new therapies can be targeted to the most appropriate patients. Current classification of myeloproliferative neoplasms is hampered by disease heterogeneity within, and clinical overlap between, subtypes. A genomic classification has the virtue of identifying patients with shared causative biologic factors, is stable over time, and does not rely on blood-count thresholds for assigning particular disease labels.

Of the eight subgroups of myeloproliferative neoplasms identified, the subgroup with *TP53* mutations was genomically unstable and had poor outcomes; this same subgroup, with similar clinical implications, has been identified in acute myeloid leukemia and other hematologic cancers.[32,35] Likewise, the subgroup of myeloproliferative neoplasms with mutations in genes regulating chromatin and RNA splicing is mirrored in both the myelodysplastic syndrome[34] and acute myeloid leukemia.[32] Patients with myeloproliferative neoplasms in this group typically had myelofibrosis, although some had essential thrombocythemia or polycythemia vera, and shared a relatively poor prognosis (as seen in patients with the myelodysplastic syndrome or acute myeloid leukemia). This raises the possibility that these driver mutations define a myeloid cancer in older patients that transcends traditional diagnostic categories.

Our model accurately identified a minority of patients with chronic-phase myeloproliferative neoplasms who were at substantial risk for disease progression. Such patients could be considered for clinical trials of new therapeutic agents, since they are the most likely to benefit and the trials would be more efficient if higher-risk patients are preferentially enrolled. Our model also

accurately identified the majority of patients with chronic-phase disease who seemingly had a benign outlook at diagnosis. In such patients, experimental therapy would be unnecessary, and a conservative treatment strategy that is based on cytoreduction and reduction of vascular risk will suffice to give long-term, event-free survival. Myeloproliferative neoplasms continue to evolve, however, and it would be informative to evaluate the opportunities offered by serial genomic profiling to update treatment choices if high-risk genomic changes emerge or if therapy drives further evolution.

Comprehensive gene sequencing of patients with blood cancers is becoming increasingly accessible and routine. The integration of clinical data with diagnostic genome profiling may provide prognostic predictions that are personally tailored to individual patients. Regarding patients with myeloproliferative neoplasms, such information will empower the clinician and support complex decisions around the choice and intensity of therapy, recruitment into clinical trials, and long-term clinical outlook.

Supported by funding from the Wellcome Trust (including a fellowship to Dr. Campbell), the Wellcome–MRC Stem Cell Institute, the National Institute for Health Research Cambridge Biomedical Research Centre, Cancer Research UK (including a fellowship to Dr. Nangalia), Bloodwise (including a fellowship to Dr. Grinfeld), the Kay Kendall Leukaemia Fund (including a fellowship to Dr. Grinfeld), the Leukemia and Lymphoma Society, the European Hematology Association (to Dr. Nangalia), the Li Ka Shing Foundation (to Dr. Wedge), and the Medical Research Council, by a grant (1005) from Associazione Italiana per la Ricerca sul Cancro (to Drs. Vannucchi and Guglielmelli), and by a grant (GR-2011-02352109) from Progetto Ministero della Salute (to Dr. Guglielmelli).

Disclosure forms provided by the authors are available with the full text of this article at NEJM.org.

We thank the members of the Cambridge Blood and Stem Cell Biobank (Cambridge) and the Cancer Genome Project laboratory (Hinxton) for technical assistance; the clinicians and staff of the centers who assisted with the Primary Thrombocythaemia 1 (PT1) studies and vorinostat trials (see the Supplementary Appendix); and all the patients who participated in this study.

APPENDIX

The authors' full names and academic degrees are as follows: Jacob Grinfeld, M.B., Ch.B., Jyoti Nangalia, Ph.D., E. Joanna Baxter, Ph.D., David C. Wedge, Ph.D., Nicos Angelopoulos, Ph.D., Robert Cantrill, Ph.D., Anna L. Godfrey, Ph.D., Elli Papaemmanuil, Ph.D., Gunes Gundem, Ph.D., Cathy MacLean, M.Sc., Julia Cook, B.Sc., Laura O'Neil, B.Sc., Sarah O'Meara, B.Sc., Jon W. Teague, B.Sc., Adam P. Butler, M.Sc., Charlie E. Massie, Ph.D., Nicholas Williams, Ph.D., Francesca L. Nice, Ph.D., Christen L. Andersen, Ph.D., Hans C. Hasselbalch, D.M.Sc., Paola Guglielmelli, Ph.D., Mary F. McMullin, M.D., Alessandro M. Vannucchi, M.D., Claire N. Harrison, D.M., Moritz Gerstung, Ph.D., Anthony R. Green, Ph.D., and Peter J. Campbell, Ph.D.

The authors' affiliations are as follows: the Wellcome–MRC Cambridge Stem Cell Institute and Cambridge Institute for Medical Research (J.G., C.E.M., F.L.N., A.R.G., P.J.C.), the Department of Haematology, University of Cambridge (J.G., E.J.B., C.M., J.C., C.E.M., F.L.N., A.R.G.), and the Department of Haematology, Cambridge University Hospitals NHS Foundation Trust (J.G., E.J.B., A.L.G., C.M., J.C., A.R.G.), Cambridge, Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus (J.N., D.C.W., N.A., E.P., G.G., L.O., S.O., J.W.T., A.P.B., N.W., P.J.C.), and the European Molecular Biology Laboratory, European Bioinformatics Institute (R.C., M.G.), Hinxton, Big Data Institute, University of Oxford, Oxford (D.C.W.), the Department of Haematology, Queen's University Belfast, Belfast (M.F.M.), and the Department of Haematology, Guy's and St. Thomas' NHS Foundation Trust, London (C.N.H.) — all in the United Kingdom; the Center for Molecular Oncology and the Department of Epidemiology and Biostatistics, Memorial Sloan Kettering Cancer Center, New York (E.P., G.G.); the Department of Hematology, Zealand University Hospital, Roskilde, and the University of Copenhagen, Copenhagen (C.L.A., H.C.H.); and the Department of Experimental and Clinical Medicine, Center of Research and Innovation of Myeloproliferative Neoplasms, Azienda Ospedaliera Universitaria Careggi, University of Florence, Florence, Italy (P.G., A.M.V.).

REFERENCES

1. Dameshek W. Some speculations on the myeloproliferative syndromes. Blood 1951;6:372-5.
2. Arber DA, Orazi A, Hasserjian R, et al. The 2016 revision to the World Health Organization classification of myeloid neoplasms and acute leukemia. Blood 2016;127:2391-405.
3. Harrison CN, Butt N, Campbell P, et al. Modification of British Committee for Standards in Haematology diagnostic criteria for essential thrombocythaemia. Br J Haematol 2014;167:421-3.
4. McMullin MF, Reilly JT, Campbell P, et al. Amendment to the guideline for diagnosis and investigation of polycythaemia/erythrocytosis. Br J Haematol 2007;138:821-2.
5. Reilly JT, McMullin MF, Beer PA, et al. Use of JAK inhibitors in the management of myelofibrosis: a revision of the British Committee for Standards in Haematology Guidelines for Investigation and Management of Myelofibrosis 2012. Br J Haematol 2014;167:418-20.
6. Wilkins BS, Erber WN, Bareford D, et al. Bone marrow pathology in essential thrombocythemia: interobserver reliability and utility for identifying disease subtypes. Blood 2008;111:60-70.
7. Barbui T, Thiele J, Vannucchi AM, Tefferi A. Rethinking the diagnostic criteria of polycythemia vera. Leukemia 2014;28:1191-5.
8. Nangalia J, Massie CE, Baxter EJ, et al. Somatic *CALR* mutations in myeloproliferative neoplasms with nonmutated JAK2. N Engl J Med 2013;369:2391-405.
9. Lundberg P, Karow A, Nienhold R, et al. Clonal evolution and clinical correlates of somatic mutations in myeloproliferative neoplasms. Blood 2014;123:2220-8.
10. Vannucchi AM, Lasho TL, Guglielmelli P, et al. Mutations and prognosis in primary myelofibrosis. Leukemia 2013;27:1861-9.
11. Tefferi A, Lasho TL, Guglielmelli P, et al. Targeted deep sequencing in polycythemia vera and essential thrombocythemia. Blood Adv 2016;1:21-30.
12. Tefferi A, Vannucchi AM. Genetic risk assessment in myeloproliferative neoplasms. Mayo Clin Proc 2017;92:1283-90.
13. Ortmann CA, Kent DG, Nangalia J, et al. Effect of mutation order on myeloproliferative neoplasms. N Engl J Med 2015;372:601-12.

14. Nangalia J, Nice FL, Wedge DC, et al. DNMT3A mutations occur early or late in patients with myeloproliferative neoplasms and mutation order influences phenotype. Haematologica 2015;100(11): e438-42.
15. Taylor J, Xiao W, Abdel-Wahab O. Diagnosis and classification of hematologic malignancies on the basis of genetics. Blood 2017;130:410-23.
16. Gerstung M, Papaemmanuil E, Martincorena I, et al. Precision oncology for acute myeloid leukemia using a knowledge bank approach. Nat Genet 2017;49: 332-40.
17. Passamonti F, Giorgino T, Mora B, et al. A clinical-molecular prognostic model to predict survival in patients with post polycythemia vera and post essential thrombocythemia myelofibrosis. Leukemia 2017; 31:2726-31.
18. Jones AV, Chase A, Silver RT, et al. JAK2 haplotype is a major risk factor for the development of myeloproliferative neoplasms. Nat Genet 2009;41:446-9.
19. Tapper W, Jones AV, Kralovics R, et al. Genetic variation at MECOM, TERT, JAK2 and HBS1L-MYB predisposes to myeloproliferative neoplasms. Nat Commun 2015;6: 6691.
20. van der Harst P, Zhang W, Mateo Leach I, et al. Seventy-five genetic loci influencing the human red blood cell. Nature 2012;492:369-75.
21. Ruark E, Snape K, Humburg P, et al. Mosaic PPM1D mutations are associated with predisposition to breast and ovarian cancer. Nature 2013;493:406-10.
22. Zhang L, Chen LH, Wan H, et al. Exome sequencing identifies somatic gain-of-function PPM1D mutations in brainstem gliomas. Nat Genet 2014;46:726-30.
23. Chen C, Liu Y, Rappaport AR, et al. MLL3 is a haploinsufficient 7q tumor suppressor in acute myeloid leukemia. Cancer Cell 2014;25:652-65.
24. Cabagnols X, Favale F, Pasquier F, et al. Presence of atypical thrombopoietin receptor (MPL) mutations in triple-negative essential thrombocythemia patients. Blood 2016;127:333-42.
25. Milosevic Feenstra JD, Nivarthi H, Gisslinger H, et al. Whole-exome sequencing identifies novel MPL and JAK2 mutations in triple-negative myeloproliferative neoplasms. Blood 2016;127:325-32.
26. Bercovich D, Ganmore I, Scott LM, et al. Mutations of JAK2 in acute lymphoblastic leukaemias associated with Down's syndrome. Lancet 2008;372:1484-92.
27. Wu Q-Y, Guo H-Y, Li F, Li Z-Y, Zeng L-Y, Xu K-L. Disruption of E627 and R683 interaction is responsible for B-cell acute lymphoblastic leukemia caused by JAK2 R683G(S) mutations. Leuk Lymphoma 2013;54:2693-700.
28. Kearney L, Gonzalez De Castro D, Yeung J, et al. Specific JAK2 mutation (JAK2R683) and multiple gene deletions in Down syndrome acute lymphoblastic leukemia. Blood 2009;113:646-8.
29. Marty C, Saint-Martin C, Pecquet C, et al. Germ-line JAK2 mutations in the kinase domain are responsible for hereditary thrombocytosis and are resistant to JAK2 and HSP90 inhibitors. Blood 2014; 123:1372-83.
30. Genovese G, Kähler AK, Handsaker RE, et al. Clonal hematopoiesis and blood-cancer risk inferred from blood DNA sequence. N Engl J Med 2014;371:2477-87.
31. Jaiswal S, Fontanillas P, Flannick J, et al. Age-related clonal hematopoiesis associated with adverse outcomes. N Engl J Med 2014;371:2488-98.
32. Papaemmanuil E, Gerstung M, Bullinger L, et al. Genomic classification and prognosis in acute myeloid leukemia. N Engl J Med 2016;374:2209-21.
33. Malcovati L, Papaemmanuil E, Ambaglio I, et al. Driver somatic mutations identify distinct disease entities within myeloid neoplasms with myelodysplasia. Blood 2014;124:1513-21.
34. Papaemmanuil E, Gerstung M, Malcovati L, et al. Clinical and biological implications of driver mutations in myelodysplastic syndromes. Blood 2013;122: 3616-27.
35. Stengel A, Kern W, Haferlach T, Meggendorfer M, Fasan A, Haferlach C. The impact of TP53 mutations and TP53 deletions on survival varies between AML, ALL, MDS and CLL: an analysis of 3307 cases. Leukemia 2017;31:705-11.
36. Hultcrantz M, Kristinsson SY, Andersson TM-L, et al. Patterns of survival among patients with myeloproliferative neoplasms diagnosed in Sweden from 1973 to 2008: a population-based study. J Clin Oncol 2012;30:2995-3001.
37. Al Assaf C, Van Obbergh F, Billiet J, et al. Analysis of phenotype and outcome in essential thrombocythemia with CALR or JAK2 mutations. Haematologica 2015; 100:893-7.
38. Cervantes F, Dupriez B, Pereira A, et al. New prognostic scoring system for primary myelofibrosis based on a study of the International Working Group for Myelofibrosis Research and Treatment. Blood 2009;113:2895-901.
39. Passamonti F, Cervantes F, Vannucchi AM, et al. A dynamic prognostic model to predict survival in primary myelofibrosis: a study by the IWG-MRT (International Working Group for Myeloproliferative Neoplasms Research and Treatment). Blood 2010;115:1703-8.
40. Passamonti F, Thiele J, Girodon F, et al. A prognostic model to predict survival in 867 World Health Organization-defined essential thrombocythemia at diagnosis: a study by the International Working Group on Myelofibrosis Research and Treatment. Blood 2012;120:1197-201.